Durham E-Theses

Peptide-based Molecular Motor Design

SMALL, LARA,SIOBHAN,REBECCA

How to cite:

SMALL, LARA,SIOBHAN,REBECCA (2015) *Peptide-based Molecular Motor Design*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/11408/

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.

Academic Support Office, Durham University, University Office, Old Elvet, Durham DH1 3HP
e-mail: e-theses.admin@dur.ac.uk Tel: +44 0191 334 6107
http://etheses.dur.ac.uk

Peptide-based Molecular Motor Design

Lara Siobhán Rebecca Small

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Department of Physics
Durham University
2015

Abstract

Peptide-based Molecular Motor Design

Lara Siobhán Rebecca Small

This thesis concerns the use of dimeric coiled-coil peptides as components for synthetic protein motors. Studies of the hub structures of two motor designs are described.

Firstly, I discuss experiments on the interactions between peptides designed for use in the Tumbleweed hub [1], a three-legged motor design containing three dimeric coiled-coil domains. Biophysical characterisation is carried out, including experiments to test the specificity of the interactions, which enable the peptides to be successful potential components for a stable hub structure.

Secondly, I discuss the design of another motor hub, with two coiled-coil domains, using similar peptides to those used in the Tumbleweed system. The requirements for this design to produce a progressive motor are discussed. The design requires one peptide spanning the length of the motor hub, which has residues involved in both coiled-coil domains. These two coiled coil-contributing regions are linked by a central span of residues. Inducing a conformational change in this central region, in order to change the dimensions of the hub, is investigated, with the introduction of an azobenzene moiety in its cis and trans isomeric forms, using MD simulations. The ability of various residues to affect the range of conformational states this central region occupies is also investigated. Experimental studies of one of the possible systems are outlined and analysed.

Contents

1 Introduction **2**

1.1 Synthetic Biology . . . 2

1.2 Natural Motors . . . 3

1.2.1 Myosins & Kinesins . . . 4

1.2.2 Dyneins . . . 5

1.2.3 Desirable Properties . . . 7

1.3 Synthetic Machines . . . 8

1.3.1 Mini Machines . . . 8

1.3.2 Designer DNA . . . 9

1.4 Peptide-based Nanostructures . . . 15

1.4.1 Overview . . . 15

1.4.2 Amino Acids and Peptides . . . 15

1.4.3 Coiled Coils . . . 18

1.4.4 Designing Coiled Coils . . . 29

1.4.5 Coiled-coil Nanostructures . . . 30

1.5 The Tumbleweed Motor . . . 32

1.5.1 Concept . . . 32

1.5.2 Timescales . . . 34

1.5.3 Simulations . . . 38

1.5.4 Components . . . 40

1.5.5 Properties Checklist . . . 46

1.6 Thesis Outline . . . 46

2 Theory of Experimental Techniques **48**

2.1 Production of Purified Peptides . . . 48

2.1.1 Solid Phase Peptide Synthesis (SPPS) . . . 48
2.1.2 High Performance Liquid Chromatography (HPLC) . . . 51
2.2 MALDI-ToF Mass Spectrometry . . . 52
2.3 Characterisation Techniques . . . 53
2.3.1 UV-Vis Spectroscopy . . . 54
2.3.2 Circular Dichroism (CD) . . . 55
2.3.3 Dynamic Light Scattering (DLS) . . . 57
2.3.4 Analytical Ultracentrifugation (AUC) . . . 59
2.4 Summary . . . 59

3 The Tumbleweed Hub 61
3.1 Hub Structure . . . 61
3.2 Experimental Aims . . . 61
3.3 Results and Discussion: Individual Peptides . . . 62
3.3.1 CD . . . 62
3.3.2 DLS . . . 70
3.3.3 AUC . . . 73
3.4 Results and Discussion: Disulphide-bonded Peptides . . . 74
3.4.1 Disulphide Bonding Between Peptides . . . 74
3.4.2 Proving Exclusivity of Coiled-coil Pairs . . . 75
3.4.3 Additional Data for Disulphide-bonded Peptides . . . 77
3.5 Conclusions . . . 80

4 Bar Motor - A Two-footed Design 82
4.1 Concept . . . 82
4.2 Coarse-grained Simulations . . . 84
4.2.1 Results and Discussion . . . 86
4.3 System Design . . . 89
4.4 Molecular Dynamics Simulations . . . 94
4.5 Initial MD Study . . . 96
4.5.1 Results . . . 97
4.5.2 Discussion . . . 100
4.5.3 Findings from Initial Simulation . . . 107
4.6 MD Simulations Round 2: Increasing the Rigidity of the Bar System . . . 109

4.6.1 Excluded Ideas for Round 2 . . . 109
4.6.2 Mutations of the Central Region . . . 111
4.6.3 Results and Discussion . . . 112
4.7 Conclusions . . . 119

5 Bar Motor Experiments **122**
5.1 Experimental Aims . . . 122
5.2 Results and Discussion . . . 123
5.2.1 CD Data . . . 123
5.2.2 UV Data . . . 126
5.2.3 DLS Data . . . 127
5.2.4 Discussion: ϵ2 With and Without Partners . . . 134
5.2.5 Discussion: ϵ2 vs. ϵ2CL . . . 135
5.2.6 Discussion: ϵ2CL With and Without Partners . . . 135
5.2.7 Discussion: Irradiation of ϵ2CL . . . 136
5.3 Conclusions . . . 137

6 Conclusions **138**

A Copyright Notices **140**

B Abbreviations **141**
B.1 Abbreviations . . . 141
B.2 Amino Acid Abbreviations . . . 144

C Product Masses **145**

D Experimental Methods **148**
D.1 Peptide Synthesis . . . 148
D.1.1 Standard Synthesis . . . 149
D.1.2 Non-standard Syntheses . . . 150
D.2 MALDI-ToF . . . 151
D.3 HPLC . . . 152
D.4 Disulphide Bonding . . . 152
D.5 Samples for Characterisation . . . 153
D.6 UV-Vis Measurements . . . 153

D.6.1 Concentration Determination . . . 153
D.6.2 Switching Effects . . . 154
D.7 Circular Dichroism . . . 154
D.7.1 Concentration of ϵ2CL . . . 154
D.8 DLS . . . 156
D.9 Glutathione Experiments . . . 156

E Spectra **158**
E.1 MALDI Spectra . . . 158
E.2 CD Melt Data for Peptide Mixtures . . . 164
E.3 Raw AUC Data . . . 165

F UV and CD Calculations **167**
F.1 Calculation of Concentration from UV Absorbance . . . 167
F.2 Circular Dichroism Units, Conversions and Calculations . . . 168
F.2.1 Explanation of CD Units . . . 168
F.2.2 Conversion of Raw Data to Final Units . . . 168
F.2.3 Theoretical Curves for Mixtures . . . 169

G Coarse-grained Simulations **170**

H Molecular Dynamics **172**
H.1 Residue Stability Checks For End-to-end Separation Measurements . . . 172
H.2 Vector-based Angle Calculations . . . 175

List of Figures

1.1 Directions of Synthetic Biology . . . 3
1.2 Structure of Myosin V in the Inactive State . . . 5
1.3 Stepping Mechanism of Myosin V . . . 6
1.4 Examples of Small Molecule-based Machines . . . 10
1.5 DNA Origami . . . 12
1.6 DNA Walkers . . . 14
1.7 Amino Acid Structure and Peptide Formation . . . 16
1.8 Peptide Secondary Structures . . . 19
1.9 Peptide Motifs . . . 20
1.10 Coiled-coil Oligomers . . . 23
1.11 Coiled Coil Structure . . . 25
1.12 Knobs-into-holes Packing . . . 25
1.13 Parallel and Antiparallel Structures . . . 26
1.14 Dimeric Coiled Coil Configurations . . . 30
1.15 Design of Orthogonal Coiled-coil Dimers . . . 31
1.16 Peptide Nanostructures from the Woolfson and Burkhard Groups . . . 33
1.17 Tumbleweed Concept . . . 35
1.18 Tumbleweed Models . . . 37
1.19 Tumbleweed Components . . . 42
1.20 Helical Wheel Diagrams . . . 45

2.1 Method of Solid Phase Peptide Synthesis . . . 49
2.2 Example HPLC Spectrum . . . 51
2.3 Example MALDI-ToF Spectrum . . . 53
2.4 Typical CD Spectra of Common Peptide Secondary Structures . . . 56
2.5 DLS Concept . . . 58

2.6 AUC 60
2.7 Measurable Properties of SE 60

3.1 The 'Y' Configuration of the Tumbleweed Hub 62
3.2 Single Peptide CD Spectra 63
3.3 Single Peptide CD Melt Spectra 63
3.4 CD Spectra of Designed Dimers 65
3.5 CD Spectra of the 12 Non-designed Peptide Pairings. 67
3.5 CD Spectra of the 12 Non-designed Peptide Pairings. 68
3.6 CD Melt Spectra of Designed Dimers. 69
3.7 CD Spectrum of Hub Peptide Mixture 69
3.8 CD Melt Spectrum of Hub Peptide Mixture 70
3.9 Sample of DLS Data 73
3.10 Disulphide Bond Reaction 74
3.11 Glutathione Reactions 76
3.12 Disulphide Bond Rearrangement 76
3.13 MALDI-ToF Spectra of Glutathione Experiments 77
3.14 CD Spectra of Disulphide-bonded Peptides 78
3.15 DLS Data of Disulphide-bonded Peptides 79

4.1 Bar Motor Concept 82
4.2 CG Model 87
4.3 CG Results: Scenario One 87
4.4 CG Results: Scenario Two 88
4.5 Helical Wheel Diagrams (Bar Motor Peptides) 90
4.6 Bar Motor Sequence Design 91
4.7 Azobenzene Linker 92
4.8 Effect of Azobenzene Switching on Peptide Structure 92
4.9 AGADIR Results for Test Sequences 95
4.10 Peptide Models for MD Simulations. 96
4.11 End-to-end Separation Measurements 98
4.11 End-to-end Separation Measurements 99
4.12 Angle Measurements 101
4.13 Evolution of Azobenzene-free Bar System 102

4.14 Evolution of Cis-azobenzene-containing Bar System . . . 103
4.15 Evolution of Trans-azobenzene-containing Bar System . . . 104
4.16 Stills from Azobenzene-free Bar System MD Simulation . . . 105
4.17 Stills from Cis-azobenzene-containing Bar System MD Simulation . . . 106
4.18 Stills from Trans-azobenzene-containing Bar System MD Simulation . . . 108
4.19 Ala27Phe Mutant of ϵ2CL . . . 111
4.20 MD Results: Free System . . . 112
4.21 MD Results: Cis System . . . 114
4.22 MD Results: Trans System . . . 115
4.23 MD Results: G26A Mutant . . . 116
4.24 MD Results: N29A Mutant . . . 117
4.25 MD Results: G26A-N29A Mutant . . . 118
4.26 MD Results: A27F Mutant . . . 119
4.27 Snapshots of MD Simulation Round 2. . . . 120
4.28 Cysteine Residue Separations in MD Simulations . . . 121

5.1 CD Spectra of ϵ2 and Nearest Analogues . . . 123
5.2 CD Spectrum of Designed Dimer ϵ, 6 . . . 123
5.3 CD Melt Spectra of ϵ and ϵ, 6 . . . 124
5.4 CD Spectrum of ϵ2,1,6. . . . 124
5.5 CD Spectra of ϵ2CL,1,6 and ϵ2CL,1-6. . . . 125
5.6 UV Spectra of ϵ2CL Recovery After Irradiation . . . 126
5.7 UV Spectra of the ϵ2CL-containing Samples' Recovery After Irradiation . . . 128
5.8 Sample of DLS Data . . . 129

E.1 MALDI-ToF Spectrum: ϵ . . . 158
E.2 MALDI-ToF Spectra: Hub Peptides 1 . . . 159
E.2 MALDI-ToF Spectra: Hub Peptides 2 . . . 160
E.3 MALDI-ToF Spectra: Disulphide-bonded Hub Peptides . . . 161
E.4 MALDI-ToF Spectra: ϵ2 Synthesis . . . 162
E.5 MALDI-ToF Spectrum: Crude ϵ2. . . . 163
E.6 MALDI-ToF Spectrum: ϵ2CL . . . 163
E.7 CD Melt Spectra: 12 Non-designed Pairs . . . 164
E.8 CD Melt Spectra: Disulphide-bonded Peptides. . . . 164

E.9 Raw AUC Data 1 165
E.10 Raw AUC Data 2 166

H.1 Evolution of Residue Separations (Figure 4.11b) 2 173
H.2 Evolution of Residue Separations (Figure 4.11b) 1 174
H.3 Angle Between Two Vectors 175
H.4 Construction of Vectors From Residue Co-ordinates 176

List of Tables

1.1 Tumbleweed Peptide Sequences 44

3.1 CD Melt Temperatures 66
3.2 DLS Averages: Tumbleweed Designed Pairs 72
3.3 AUC Data: Designed Pairs 74
3.4 AUC Data: Disulphide-bonded Peptides 80

4.1 MD Simulation Sequences 113

5.1 DLS Averages: Bar Motor ϵ2 Samples 131
5.2 DLS Averages: Bar Motor ϵ2CL Samples 132
5.3 DLS Averages (time spaced): Irradiated ϵ2CL 133
5.4 DLS Averages (time spaced): Irradiated ϵ2CL, 1 ,6 133
5.5 DLS Averages (time spaced): Irradiated ϵ2CL, 1-6 134

B.1 Amino Acid One- and Three-letter Codes 144

C.1 Disulphide-bonded Peptide Masses 145
C.2 Peptide Masses 146

D.1 CD Parameters 155
D.2 CD Melt Parameters 1 155
D.3 CD Melt Parameters 2 156

Declaration

The work presented here is the work of the author (LSRS) unless stated.

The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged.

Acknowledgements

I would like to thank my supervisor Beth Bromley for her support during my time in Durham. I would also like to give my thanks to our collaborators. To Martin Zuckermann, for providing coarse-grained simulation data to complement my work, and answering my questions. To Richard B. Sessions, for his assistance in all things regarding Molecular Dynamics, and for being so accommodating during my visit to Bristol. To Dek Woolfson and his group, particularly Drew and Aimee, for their help during my visits. Many thanks to Asahi, my Bromley group peer and fellow Dr Pepper enthusiast, for $\epsilon 2$ and much more. Thanks to the Department of Chemistry Mass Spectrometry Service.

My thanks go to the Cobb lab, my surrogate group, whose presence regularly proved to be more companionable than that of my peptides, for your friendship and reassurance in chemistry-based matters. In the Physics Department, I am thankful for my officemates and PhD colleagues past and present; I thank them for their company at lunchtimes, and sanity-saving board game evenings, badminton games and office catch breaks. For their programming experience I thank Ewan, Tom, Pete, and Ranga. Thanks to Lynne for the provision of many pints of squash, and for being there when I needed a chat. Thank you, Rod and Mark, for all the support, conversations, and reminding me of the importance of focussing on the present.

Thanks to my fellow Ustinovians, without whom my first year would have been completely different, and my housemates Jon, Jo, and now Ste, for putting up with me. To Sam, Sam, Hannah, Tom, Tom, Dawn, Emma, Rebecca, Fliss, and the many others whose support has always been much appreciated.

Finally, many thanks to my family, especially Bev and Ger, my ever supportive parents, who, without pressure, have always helped me wander in the right direction, and my nan Sheila and brother Thomas, for always being there when I needed them.

Chapter 1

Introduction

This thesis discusses research into the use of designed coiled-coil peptides as motor components. The introductory chapter, therefore, initially covers natural motors, the inspiration for such work, their hierarchy of structure and function, and their features that are desirable to reproduce through synthetic biology; synthetic machines that have been produced thus far, from a range of materials, are then covered lightly, before focus is brought to the previous work on the Tumbleweed, the continuation of which is a major component of this work.

1.1 Synthetic Biology

Synthetic biology can both confirm our ideas about how natural systems function, and allow us to use nature's construction rules to make new systems with designed attributes [2]. Work in the area is greatly varied, from the synthesis of novel amino acids [3], through track-walking DNA molecules [4; 5; 6; 7; 8; 9; 10; 11; 12; 13], to the synthesis of a genome which is replicated when put into a cell [14]. A summary of the pathways taken by research into synthetic biology is shown in Figure 1.1. For constructions from simple units, there is the use of naturally occurring small molecules (natural amino acids, DNA bases), or the synthesis of designed molecules, such as peptide nucleic acids (PNAs) and β-amino acids, to build designed higher-order structures. Other work starts with the use of larger structures found in nature, but moving them to new environments. An example of this is Path 5 of Figure 1.1, the path taken by the Venter Institute [15], where researchers transplanted the whole genome of one bacterium into another (they have since transplanted a full synthetic genome into bacteria cells [14], published after Figure 1.1). Our work follows Path 8, using

designed peptides for synthetic biology.

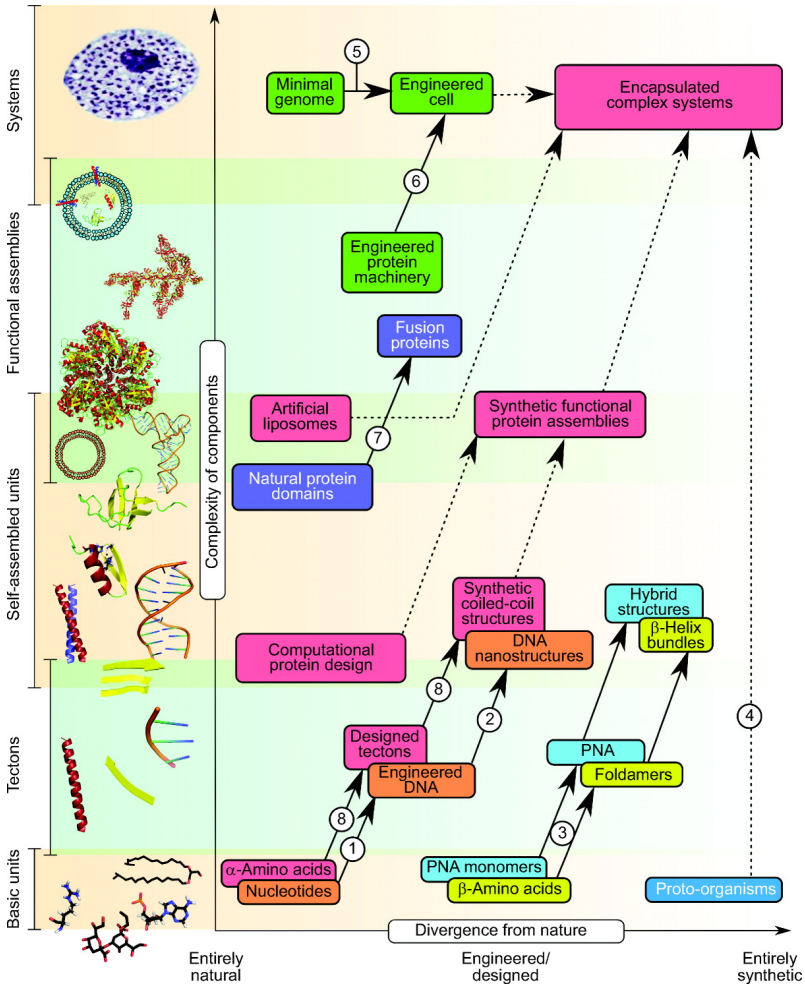


Figure 1.1: The various directions of synthetic biology. The bottom left of the diagram shows the natural units used by nature to construct systems, while the top right shows completely synthetic systems. The hub of the Tumbleweed synthetic motor [1] follows Path 8. Copyright notice in Appendix A [2].

1.2 Natural Motors

Molecular motors are prevalent in mammalian cells. Three 'families' of motors, kinesins, myosins and dyneins, use binding and hydrolysis of ATP to move, transporting cargo around the busy cellular environment along actin filaments and microtubules [16]. They are natural converters of chemical energy into mechanical work. The roles they play are extensive and diverse. For example, myosins are responsible for generating muscle contrac-

tion, kinesins are involved in mitosis (cell division), moving vesicles and mitochondria, and dyneins move cilia and flagella. There are differences between the families, but particularly when comparing myosin and kinesin with dynein. Myosins travel along actin filaments, while kinesins and dyneins both travel along microtubules (kinesins travel toward the 'plus' polar ends, to the cell edges, and dyneins travel inwards where microtubules converge at a central point), but kinesins and myosins are much more structurally alike than dyneins.

1.2.1 Myosins & Kinesins

Myosins and kinesins are both dimers, held together through coiled coil formation (a property used in the Tumbleweed and Bar motor designs), with 'heads' which bind to their respective tracks, upon nucleotide binding (myosin V and kinesin-1 have one binding site in each of their two heads), and 'tails' bound to cargo (Figure 1.2). Both of these motor proteins travel hand-over-hand (so the heads alternate between being the leading and trailing head) [17; 18; 19] (Figure 1.3), using one ATP molecule for each step, and have good levels of processivity - they can take multiple steps before detaching from their track. Kinesin-1 has a fixed step size of ∼16 nm (equating to a centre of mass change of ∼8.3 nm) per ATP hydrolysis [18; 20], and is only slowed by loads over 3pN. It does not tend to take backsteps [16]. It travels along just one protofilament of a microtubule and can take around 100 consecutive steps in vitro [16]. Myosin V, by contrast, shows significant backstepping for loads over 1pN, and also can take 'halfsteps' [16]. Myosin V's steps are ∼36 nm, with a resultant centre of mass change of ∼18 nm [21].

The stepping process of myosin V is shown in Figure 1.3. Myosin V is active when cargo is present; it is otherwise inactive (Figure 1.2), with the motor and cargo-binding domains interacting with each other in a more compact structure than the active one. In order for myosin to progress along actin filaments, ATP is needed. The ATP binds to a head domain, and in both the ATP-bound, and subsequent ADP-P_i-bound, states, it has a low affinity for actin. Once P_i is released, and only ADP is still bound, the head's affinity for actin is high, as is the nucleotide-free state after ADP is released. As shown in Figure 1.3a, with both feet bound to ADP, they are also bound to the actin filament. Once a foot releases ADP (1.3b; it is thought that strain exerted between the two heads on each other is how the trailing foot loses ADP first [21]), ATP can bind and cause the foot to lift (1.3c). Hydrolysis of ATP to ADP and P_i provides a power stroke and the foot moves forwards (1.3d). Actin binding, followed by the release of P_i, then allows the foot to more

strongly bind actin, 36 nm ahead of its previous site, and the cycle repeats (1.3a). The phosphate release (shown in Figure 1.3) and ATP hydrolysis (>250s^{-1}) are both rapid steps, while ADP release from the actin-myosin is slower (12-16s^{-1} [22]), limiting the rate of the process [22; 21]. It is thought that when both heads are bound with ADP and actin, it is the strain exerted on the two heads which causes the trailing one to release its ADP whilst preventing release from the leading head [23; 24; 25; 26], giving the motor directionality.

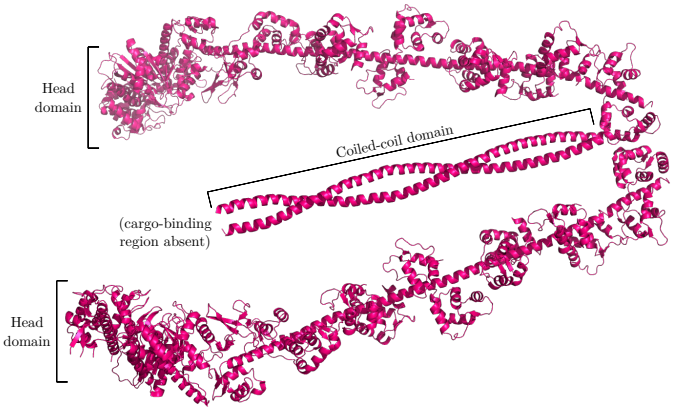


Figure 1.2: Structure of myosin V in the inactive state. Shown (in the centre) is the coiled coil which drives the protein's dimer formation, surrounded by the chains' neck and head domains. The neck domains each bind six calmodulin molecules, while the head domains bind ATP/ADP. At the other end of the coiled-coil domain, a cargo-binding domain is present in the full motor protein (not shown in this structure). Figure made in PyMol [27] from PDB file 2DFS [28].

1.2.2 Dyneins

In comparison, only a few of the dyneins found are processive, and hence of interest in terms of the properties we are trying to replicate with our synthetic system. Cytoplasmic dynein-1 has two heads of far greater complexity than those of the other families, the motor itself being around 10 times the size of kinesin-1 [16]. Its heads have more than one binding site and dynein also makes use of other proteins (accessory proteins), which appear to be important for it to function. Its step size varies more than for the other motors, depending on the load attached, and this increased complexity may make dynein more open to controlled variability in motion. Its motion is highly variable, with both stochastic

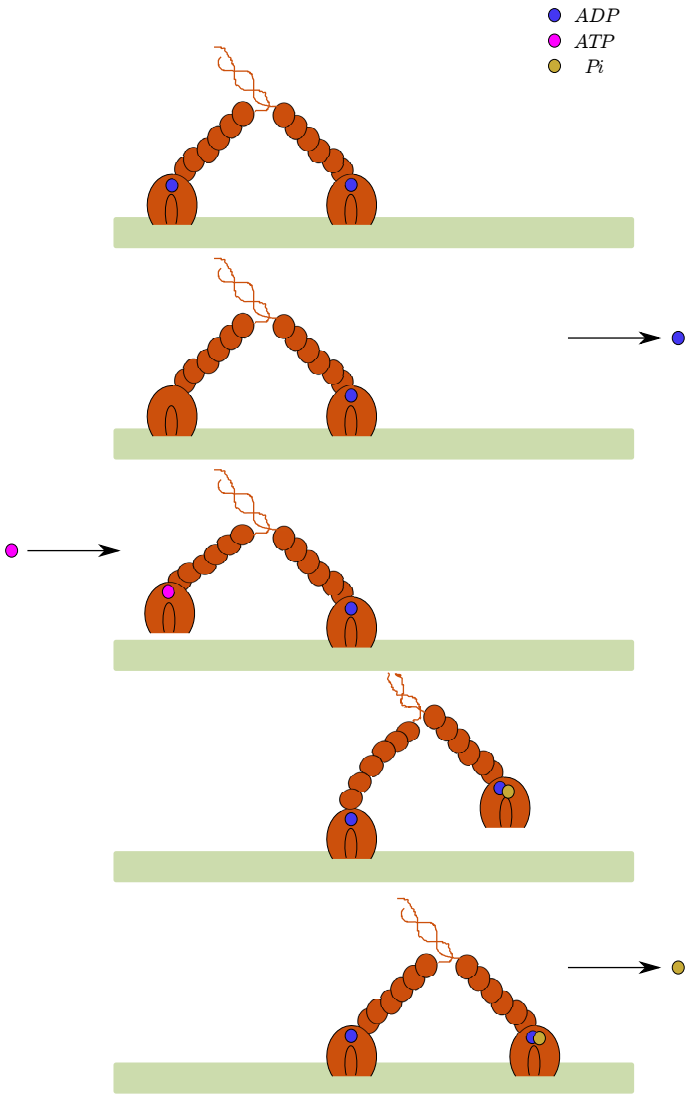


Figure 1.3: Stepping mechanism of myosin V along an actin filament. Figure adapted from [21]. While both heads are ADP-bound, they have a high affinity for actin (a). It is when ADP is released (b), and ATP is able to bind, that a lower actin affinity causes the head to release (c). ATP hydrolysis provides the power stroke, and the now ADP-plus-phosphate-containing head moves ahead of the other to its new position (d), binds, and then releases the phosphate group (e). This returns it to the high actin affinity state of ADP (a) [21].

and coordinated stepping, seemingly dependent on the separation of its motor domains [20]; when both feet are close to each other, steps seem mostly stochastic, by either foot, but as separation increases, stepping is increasingly by the back foot. Qiu et al. found that 74% of steps were temporally alternating between feet, but in 83% of the steps, the stepping foot was not spatially overtaking the other foot [20]. Cytoplasmic dynein's step size seems to be around 14 - 16 nm in 2D, while its tail (a measure similar to centre of mass change given for other motors) moves around 10 nm, and 8 nm in the direction of motion along a microtubule [20]. Unlike kinesin, it uses multiple protofilaments, and, due to their sharing a common track, it has been suggested that dynein may move to other protofilments if a kinesin is travelling along the same path. It is thought that dynein motors may work in teams to transport larger cargo [16].

1.2.3 Desirable Properties

The different motor proteins offer different models for motor design - it may be simpler to create a motor that moves like kinesin or myosin, but it could also be of interest to have the range of potential variability which dynein offers.
Some of the desirable properties of biological motors, advantageous to be reproducible in synthetic motors, are as follows:

1. Autonomous Motion: the ability of motors to move independently, without external control, taking multiple steps without an external influence added/changed per step.

2. Directional motion: the ability of motors to walk in a particular direction, either due to the track or the motor itself.

3. Processive motion: a successful motor must take multiple steps without dissociating and falling off the track.

4. Efficiency: obviously efficiency is desirable, with minimal missteps and stalling.

5. Structural stability: for the motor molecule itself, being stable enough that its parts do not spontaneously dissociate is a necessity.

6. Cargo binding ability: controllably binding/unbinding a cargo (and transporting it) would be an advantage, and a likely means of giving a motor a clear function.

7. Environmental sensitivity: an efficient motor will only carry out its task when needed, informed by interaction (signal/response) with its surroundings. The ability

to provide a fuel which produces a power stroke would be a very desirable property in a synthetic motor.

1.3 Synthetic Machines

Complementing, and inspired by, the research and growing knowledge of natural motor function and structure, is the work using this knowledge, and testing our understanding, through the construction of designed objects, either made from biological molecules, or trying to emulate natural machine functions.

'Current technologies are focused on miniaturization of functional systems with molecular machines as one of the ultimate goals'

Ariga, Mori & Hill (2012) [29]

The field of nanoscale machines is diverse. There are small 'machines', such as rotors, switches and now even walkers of some description, made from small molecules, through to larger constructions made from biological building blocks such as DNA nucleotides and amino acids. In terms of progress, work involving DNA is more advanced than peptide work; DNA Watson-Crick base pairing [30] makes it easy to design complementary strands, and research using DNA for nanostructures has continued to thrive, with increasingly complex structures being produced. This advantage in terms of ease of design also provides disadvantages, as the complexity afforded is limited. The possible interactions between amino acids are far greater and more complex; peptides form a variety of secondary, and higher order, structures. In the grand scheme of things, peptide work also has the advantage over DNA that natural biological motors are proteins, and hence advances made should also increase our understanding of the natural world. There are also natural machines such as the ribosome which consists of both RNA and proteins [31]. Here are a few examples of what has been achieved with small molecules and DNA in terms of molecular structures/machines. A more detailed introduction to peptide structure, and peptide nanostructures is in the next section.

1.3.1 Mini Machines

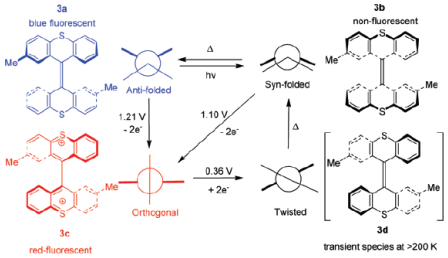
Small molecule machines vary greatly, and just a few are discussed here to give an impression of the vast landscape of nanomachines and the desirable properties that can be

realised. Five examples are shown in Figure 1.4. Some are fairly simple, such as molecular switches (1.4a) and molecular shuttles (1.4b). Some more complex examples include the nanocar (1.4c), which tries to replicate motion by emulating macroscopic motion, and the Leigh group's walker (1.4d) [32] and peptide synthesiser (1.4e) [33].

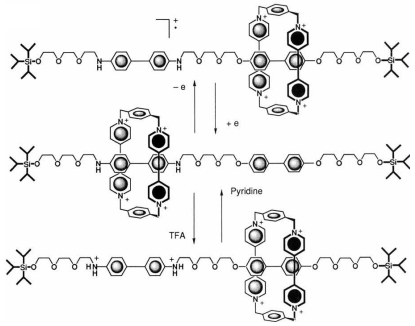
The bisthioxanthylidene switch (1.4a) controllably switches conformation, due to thermal, electrochemical and photochemical changes [34; 35], and these conformations correspond to red, blue and non-fluorescent states. In the rotaxane shuttle (1.4b), the ring has a preference for the left benzidine unit on the axle (84% of rings are in the left position in CD_3CN solution at equilibrium), but can be switched to the right-side biphenol group by electrochemical oxidation/protonation [36]. A rotaxane ring-axle structure also forms the basis of the synthetic peptide synthesiser (1.4e); the ring travels along an axle, performing native chemical ligations (NCL) with amino acid phenolate esters placed along it, forming a peptide chain of defined sequence [33]. In the Leigh walker (1.4d), metal-complex feet bind to the anchorages, with one (Pt(II)) foot remaining bound at the central anchorage at all times, and a Pd(II) foot switching between the two other sites, induced by protonation/deprotonation and thermal activation [32]. The nanocar takes a slightly different design direction, that of miniaturization; it has fullerene wheels that roll like motorcar tyres, as opposed to slide, to perform translational motion [37]. The version in Figure 1.4c has a chassis sufficiently flexible to overcome one-atom-high gold hurdles [38], but no means of carrying a cargo. Between these small machines, most of the desirable properties listed in 1.2.3 have been demonstrated. However, none of these have the full set, and while this would be an achievement, it would not greatly assist our understanding of biological motors due to the large differences in composition.

1.3.2 Designer DNA

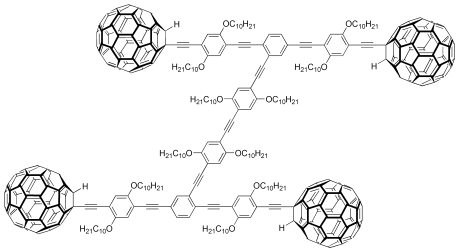
A step closer to natural motors than small molecules is DNA. DNA design takes advantage of the strength of interaction between Watson-Crick base pairs (cytosine (C) - guanine (G) and adenine (A) - thymine (T)) [30]. In nature, the ratios of C:G and A:T are close to 1:1 [39; 40; 41; 42]; other base pairings are less stable, with the alternative hetero pairings less stable than homo pairs [43]. The structures of these two pyramidine-purine (C-G, T-A) pairings have comparable sizes, enabling a regular double helix with a constant diameter to be formed [44; 43]. This almost-exclusive pairing of A with T and C with G means that creating complementary strands (and making others uncomplementary) is straight-



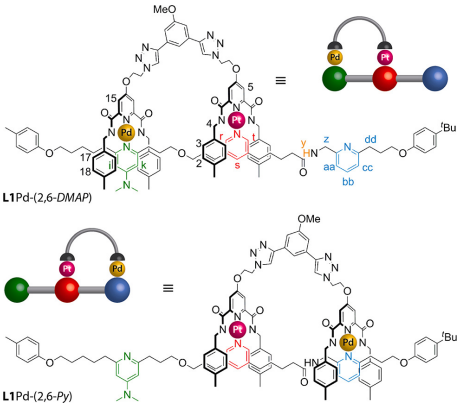
(a) Bisthioxanthylidene: thermal, electrochemical and photochemical switches induce conformational changes.



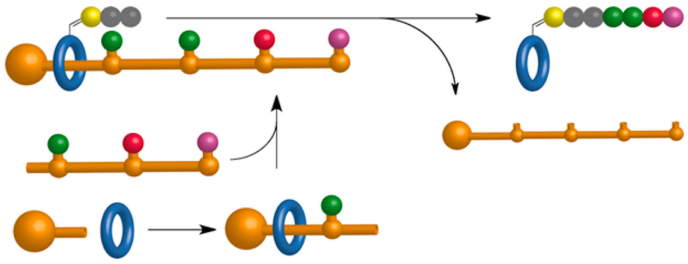
(b) Molecular shuttle: a rotaxane shifts position with electrochemical oxidation/protonation.



(c) One of a family of nanovehicles: the 'nanocar' can move translationally using fullerene wheels which roll like car tyres.



(d) Two-footed 'walker' on a three anchorage track: thermal activation and protonation/deprotonation drive the motion of one foot between two positions.



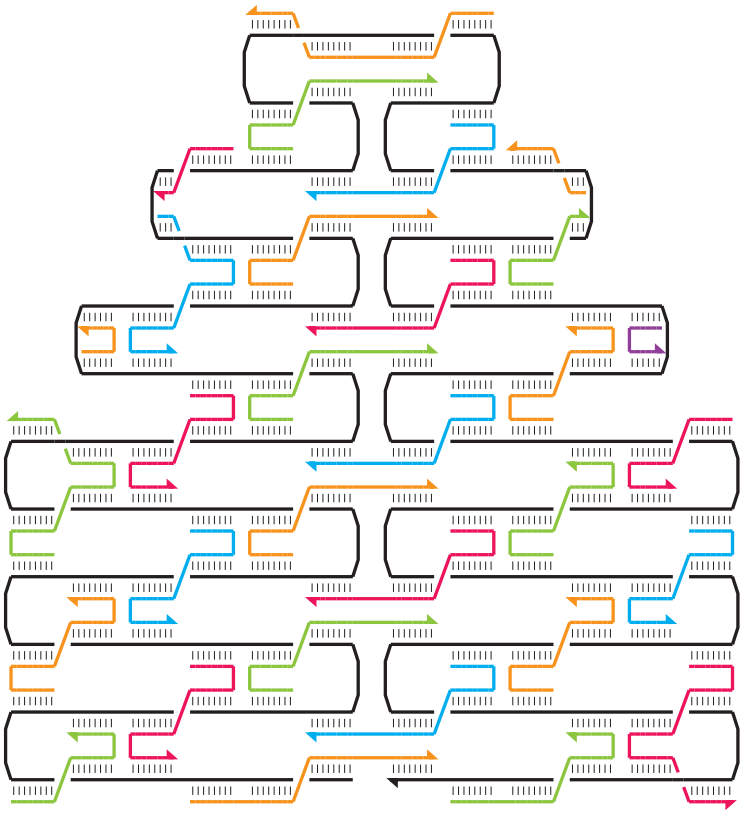
(e) Molecular peptide synthesiser: NCL reactions occur at each step, allowing the rotaxane to progress and build up a peptide chain.

Figure 1.4: Examples of small molecule-based machines. Copyright notices in Appendix A [34; 36; 38; 32; 33].

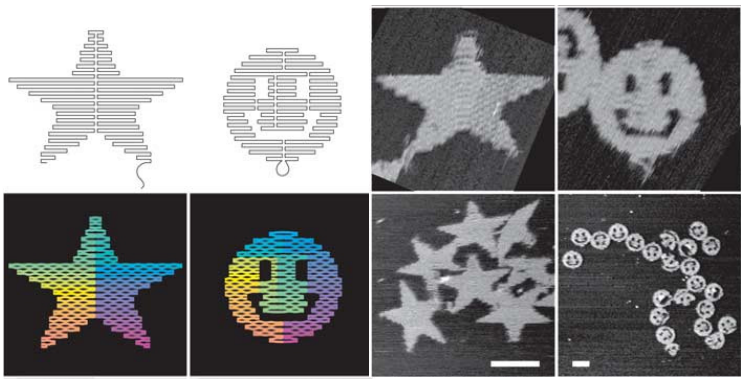
forward.

DNA structure is driven by finding an energetically favourable state, so designing sequences which make the desired structures highly preferable allows DNA to be used as a nanomaterial. If the formation of a cube is desired, then the formation of a cube should, by design, maximise the number of base pairs satisfied in the designed sequences. To cause a change in state, such as a bonding change, the new state has to be more desirable, either due to a reduction in the current state's stability or by opening up a more desirable state to the system. In many designs, changes in structure are made possible by including 'toeholds' - single stranded stretches of DNA at the end of double strands. If a toehold-containing strand (the target strand) is bound to a complementary strand by all bases except for the toehold bases, then a suitable incoming strand can hybridise to the toehold, and a displacement reaction can take place to replace the first complementary strand with the new strand. The toehold can either be present in the initial system, and an incoming strand create more base pairs than the initial arrangement, or be created by hydrolysis, induced by nicking enzymes, which are only activated by the first complementary strand binding to the target strand (thus completing a recognition sequence for the enzyme). In this case, the section cleaved by the hydrolysis is just a few nucleotides, which can spontaneously dissociate from the double strand [45].

An important development in the field of DNA nanotechnology was that of DNA origami. In this now famous 2006 paper, Paul Rothemund describes the technique of using one long DNA strand and many smaller complementary strands ('staples') to form a variety of shapes, and demonstrates several, including a star and smiley face, to show the generality of the concept. For a given shape, a design is made comprising of rows of helices, and crosslinks connecting the rows. A single unbroken strand is then drawn through every helix, crosslinking them in different positions to the first set of crosslinks. The original crosslink markings are replaced with short DNA strands, which then link two helices via base pairing; wherever two of these short strands back into each other, they are both modified so that each links one of its own helices with one of the other staples' (see Figure 1.5). A computer program is used to find suitable sequences and the shapes formed just by using an excess of staples with long strands and annealation for under two hours (95°C to 20°C). The paper has been cited over 1000 times [46], the method is now regularly used for many synthetic nanostructures, including as the foundation for DNA walker tracks [12; 13], and a 3D analogue of the technique has been successfully implemented [47].



(a) One long DNA strand (black) forms half of each helix in the shape, and binding with complementary smaller staple strands (coloured) guides the formation of a desired shape.



(b) Two of the origami designs. Top left: Folding paths. Bottom left: Bending of helices, showing where they touch at crossovers and bend apart away from them. First to last DNA bases indicated by transition from red through yellow, green and blue to purple. Top right: 165 nm x 165 nm AFM images. Bottom right: AFM images with 100 nm scale bars shown.

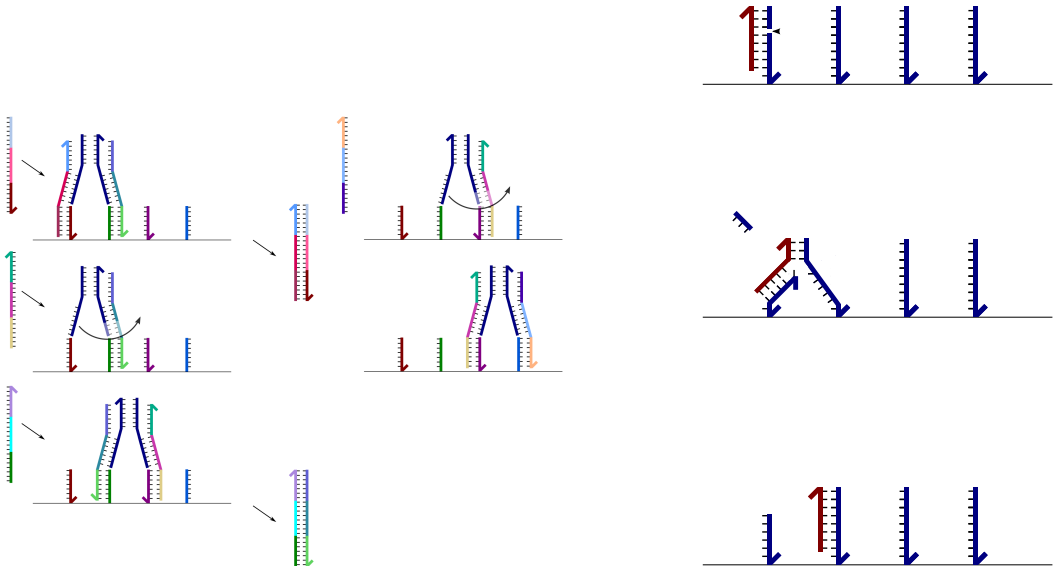
Figure 1.5: The concept of DNA origami, and AFM images showing proof of concept. Copyright notice in Appendix A.

Dynamic Structures

Having controllable functions for these nanostructures is desirable. The Turberfield group in Oxford has some nice examples of nanostructures with easily visible uses. They developed a tetrahedron formed from four DNA strands (each forming parts of three tetrahedron sides) [48], and have since modified this design, to allow a conformational change (using a 'fuel' strand that comes in and opens a hairpin, extending one edge) which is reversible (using an 'antifuel' strand) [49], enclosed molecules within tetrahedra through covalent attachment prior to tetrahedron formation [50], and non-covalently, through DNA recognition binding post-tetrahedron formation [51]. Molecules have been released from the tetrahedra by degrading the enclosure [51], and tetrahedra have been delivered to human embryonic kidney cells cytoplasm, and remained intact for at least two days [52], demonstrating characteristics that are required in potential drug delivery systems.

DNA has also been used to create several walkers [6; 7; 4; 8; 5; 11; 9; 10; 13]. These walkers use a couple of common ideas in different ways and combinations. The main methods of binding between a track and a walker are by the two having complementary strands that hybridise, or by connecting the two via a third strand, which hybridises with part of a track anchorage, and part of a motor strand [6; 7]. These are made processive by using other strands to displace a linking strand [6; 7], such as in the motor shown in Figure 1.6a, or by destruction of part of the track [4] so as to make movement favoured (an example of which is shown in Figure 1.6b). Some of these motors require external control, such as the addition of different DNA strands [6; 7], but autonomous motion, without the necessity of external additions, is desirable. Hydrolysis of DNA [4; 5], ATP hydrolysis [8] and again, DNA hybridisation [9; 11; 10], as the fuel for motion have been used. The Turberfield group have extended their single strand autonomous walker work to demonstate the ability to control the direction of a motor given a choice of paths. They have done this using blocking strands at junctions, and by controlling the unblocked path through a) the addition of unblocking strands, and b) the inclusion of a region in the motor strand itself which catalyses an otherwise hindered block-unblock hybridisation [13].

Some of these walkers are much closer to the natural motors discussed in Section 1.2 than the small machines (Section 1.3.1), and alongside, or combined with, the static DNA structures, may be suitable for a number of applications. Our aim, however, in addition to creating structures which can be given applications in the future, is to further our understanding of the interplay of interactions in natural motors and machines. For this we



(a) Shin & Pierce's two-strand walker [6] binds each foot to an anchorage with a different 'linking' strand, and moves the motor along (once both feet are bound to anchorages) by adding in both a complementary 'unlinking' strand, which displaces the bonds between one linking strand and its motor foot and anchorage, and a new linking strand that can bind the newly-released foot to the next free anchorage. The system tackles the issue of backstepping by having four differently sequenced anchorages, instead of two, so that the new linking strand added in is highly biased to binding the foot to the forward anchorage (provided that dimensional constraints are sufficient that the motor can't reach back four steps).

(b) Bath, Green & Turberfield's single-strand DNA motor and DNA track [4] has anchorages which are all the same. The track is therefore less complex than that of Shin & Pierce, but to impart directionality, a restriction enzyme cuts part of the anchorage when the motor/cargo is bound, destroying the track and hence preventing backstepping (the free energy change to do so would be large). The DNA cargo moves to the next site, where all of the base pairs can bind again, by reaching over to the next anchorage strand where the recently-created unbasepaired region of the motor strand is also satisfied. All of the still-bound nucleotides on the first anchorage then need to dissociate at the same time in order for the move to take place.

Figure 1.6: Two popular methods of directional DNA walkers.

need to use the same starting materials, amino acids.

1.4 Peptide-based Nanostructures

1.4.1 Overview

The wider range of amino acid building blocks and their interactions makes peptide structures more difficult to predict than DNA constructs, but means there are greater possibilities open for *de novo* design, and there is a growing bank of knowledge on designing desired peptide structures.

De novo peptide design, like many biophysical fields, has motivations in medical solutions. Several groups have taken natural peptide sequences, modified them [53], or designed their own [54; 55; 56; 57; 58; 59; 60; 61] to form hydrogels [62]. Some of these β-structured hydrogels have had properties such as antibacterial activity [63], no immune response [53], and the ability to support cell growth [64]. Fmoc dipeptide-based hydrogels have shown the pH of gelling and cell support to be amino acid dependent, and that one simple system of an Fmoc-dipeptide (diphenylalanine) and tripeptide (arginine-glycine-aspartate) can bind integrin proteins and support cells which need to adhere to the scaffold to grow [65]. Several fibrous peptides have also been designed, and form gels and scaffolds with α-helical [66; 59], coiled-coil [56; 57; 58] and collagen/collagen-like [61; 67; 68; 60] structures. Peptide amphiphiles, usually with both a polar peptide and a non-polar aliphatic region [62], have been shown to have some impressive properties, including the ability to help spinal cord injury recovery in mice [69; 70], and prevent cancer cell growth (by disruption of a protein-protein interaction) [71]. Antiviral/antibacterial peptide nanotubes have also been produced [72; 73; 74], while coiled coils/coiled coil-phospholipid composites have been used to create nanoparticles with antigens attached [75; 76].

As this project principally concerns coiled-coil peptide domains, this section will discuss the formation of peptides and coiled coils, followed by developments in nanostructures, with particular focus on those formed from coiled coils.

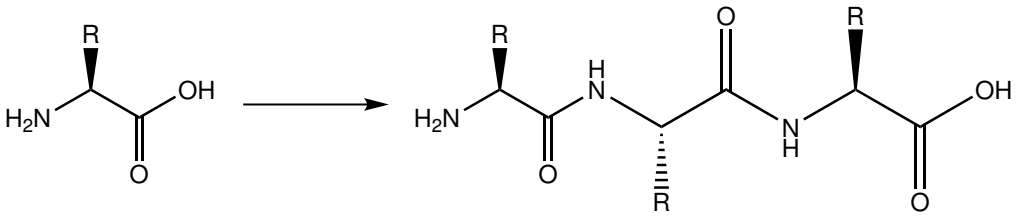
1.4.2 Amino Acids and Peptides

Formation

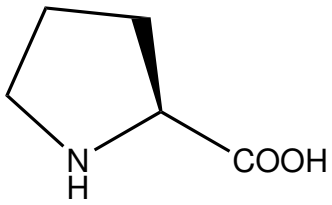
Peptides are chains of amino acids, of which there are 20 in the canonical alphabet coded for directly in DNA [77] (these are listed with their one- and three-letter abbreviations in

Appendix B.2, which will be used throughout this thesis). Eleven of these are synthesised by the human body, and the other nine have to be ingested ('essential' amino acids). There are many others found naturally, including a few only found in prokaryotes, or coded for only in the presence of other specific molecules, and many that are the result of post-translational modifications to canonical amino acids [77]. Outside the natural world, there are unnatural amino acids, such as those being chemically synthesised, intended to possess more desirable pharmacological properties than their natural counterparts, such as a resistance to enzymatic degradation [78]. The peptides in this project only use some of the standard 20 amino acids.

To form peptides, amino acids undergo condensation reactions to form a covalent bond between the carbon atom of one amino acid's carboxyl (-COOH) group and the nitrogen atom of the next molecule's amino group ($-NH_2$), as shown in Figure 1.7a. The variation



(a) The general structure of an amino acid ($H_2NCHRCOOH$) is shown, followed by the structure of a tripeptide, formed from the reaction between amino acids to form peptide bonds. A water molecule is released for every additional amino acid bound. The end of the peptide with its $-NH_2$ group intact is the N-terminus, and the -COOH end is the C-terminus. The R group is the part of the structure which varies between different amino acids.



(b) The only exception to the general structure shown in (a) in the standard 20 amino acid alphabet is proline, with its ring structure replacing the (usually separate) $-NH_2$ and R groups.

Figure 1.7: Amino Acid Structure and Peptide Formation.

in side chains (represented by the 'R' group) gives the amino acids different properties. The amino acids are often grouped into those which are hydrophobic (F, A, M, I, L, Y, V,

W) and those which are polar (D, E, H, K, N, Q, R, S, T) [62]. Glycine, proline (Figure 1.7b) and cysteine are less easily grouped. Glycine is the simplest (its R group is just a hydrogen atom), and only achiral proteinogenic amino acid. Cysteine, due to the sulphur atom in its side chain, can form disulphide bonds, a property we make use of in this project. Proline does not have the same general structure as the other 19 (Figure 1.7); its ring structure gives it reduced conformational flexibility. The presence of an aromatic moiety in phenylalanine (F), tryptophan (W) and tyrosine (Y) means that they absorb UV light. There are also negatively (D, E) and positively (R, H, K) charged amino acids, another property regularly used in coiled-coil design. The sequence of amino acids forming a peptide is known as the peptide's primary structure.

Secondary Structure

The varying properties of different amino acids allow different peptide secondary structures to form. The two most common peptide secondary structures are the α-helix and the β-sheet [79], and then there are other (less structured) structures such as loops, turns and random coils.

An α-helix (Figure 1.8, a&b) has 3.6 residues (amino acids) per turn, and is held in shape by hydrogen bonds between the ith and (i+4)th residues [79]. Helices vary from four residues to over 40 residues and are most commonly found on the surfaces of proteins. Consequently, α-helix sequences often have hydrophobic residues every three or four residues, so that one side of the helix is hydrophobic and the other is polar; if on a protein surface, the helix will have its hydrophobic surface buried inwards and the polar surface facing outwards towards the solvent.

The amino acids, unsurprisingly, have different helical propensities. Results vary from study to study, depending on the method and model used, due to the different intertwined interactions of the system. It is often impossible to completely separate out different effects, but there are common features of all of the studies. Alanine is consistently found to have a high helical propensity [80; 81; 82]. The restriction of its conformational space due to its methyl side chain appears to be the reasoning behind this [80]; glycine, lacking in such a side chain, is one of the most destabilising amino acids in α-helices, due to its great conformational flexibility [81; 80]. Proline is consistently, out of the twenty most common amino acids, least suited to an α-helical environment [82; 81; 80], and results in a kink or bend in the helix [83].

The β-sheet (Figure 1.8, c&d) is another common secondary structure. β-strands, generally between five and ten residues long [79], hydrogen bond to each other to form β-sheets. These can have mixed, parallel or antiparallel forms, defined as whether all of the N- to C-termini of the strands run in the same (parallel) or alternating (antiparallel) directions.

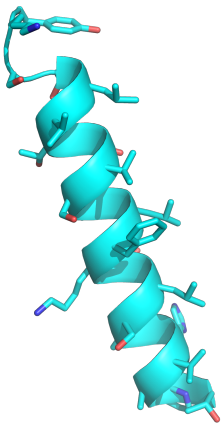
Tertiary and Quaternary Structure

From these secondary structures, motifs can be formed - intermediates between the secondary structures and full blown proteins. They are small collections of secondary structures joined together in a combination found regularly in proteins [79]. Some examples (shown in Figure 1.9) are the zinc finger (where, classically, two cysteine and two histidine residues bind a zinc atom, and the residues between the second cysteine and first histidine are known as the finger [79]), the helix-turn-helix (found in many DNA binding proteins such as TrpR [86], PurR [87], CAP [88], cro repressor [89; 90] and λ repressor [91]), collagens (the most common proteins in humans, non-exhaustively, found in the skin, tendons and bones [92]; three parallel (Gly-X-Y)-repeating chains supercoiled together, where X and Y are commonly proline and hydroxyproline [79]), and the coiled coil (a motif that can form by the wrapping together of more than one alpha helix into a more stable structure [79], discussed in Section 1.4.3 onwards). There may be more than one motif in a peptide chain, and the way that the whole peptide chain folds up into a more globular form is its tertiary structure. Many proteins are made up of several chains, and the interaction between these already folded chains to give the final 3D protein structure is the quaternary structure.

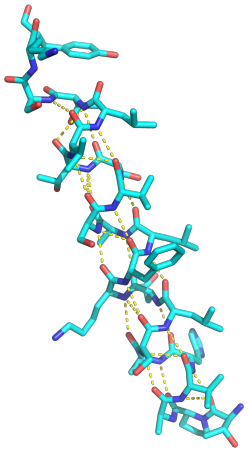
1.4.3 Coiled Coils

Coiled coils are well suited to the production of small self-assembling units, making them a good candidate for synthetic design.

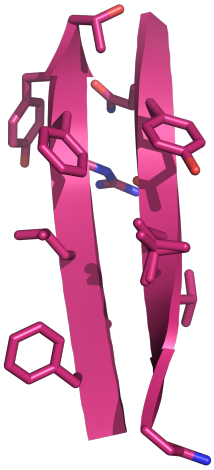
Coiled coils have been a known structure since 1953 [97; 98], proposed as a solution when descriptions of α-keratin as an α-helix led to inaccurate density calculations, and measurements which did not correspond to its diffraction pattern. Crick [97] suggested that if helices wound together, it would slightly reduce the number of residues per turn, and with knobs (side chains) into holes (spaces between side chains) packing (Figure 1.12), they would be able to remain in such a conformation; the resultant density from this structure was much closer than that of an α-helix to the α-keratin observed density. The energy



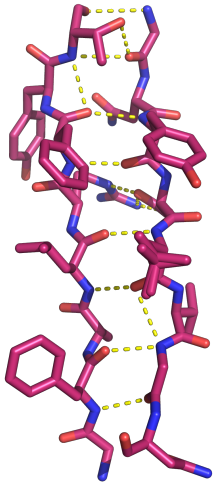
(a) The α-helix is commonly represented as a spiral (as shown here, with the peptide side chains also shown) to represent the helical backbone.



(b) It is far more difficult to see the α-helical backbone in a representation showing both the backbone and sidechains. The hydrogen bonds stabilising the structure are also shown here.

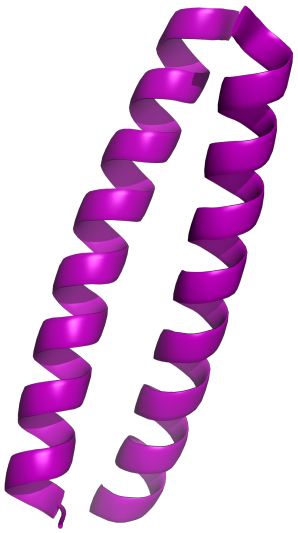


(c) The β-strands of a β-sheet are commonly represented by arrows, pointing from the N- to C-terminus (side chains also shown here).

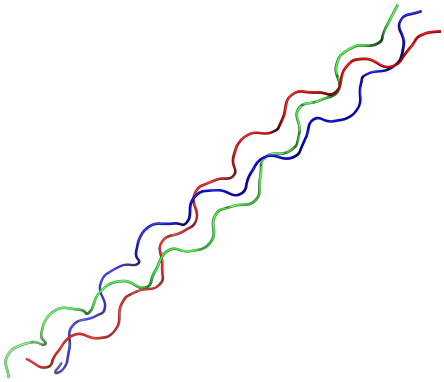


(d) The β-sheet is held together by a network of hydrogen bonds.

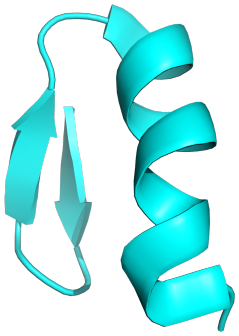
Figure 1.8: Peptide secondary structures. The α-helix and β-sheet are the two most common secondary structures. Figures made with PyMol [27] using PDB files 1GZX (residues 118-140) [84] of human haemoglobin, and 1GWE (residues 116-123 and 126-133) [85], of catalase from Micrococcus lysodeikticus.



(a) Dimeric coiled coil: peptide corresponding to the leucine zipper of yeast (Saccharomyces cerevisiae) transcriptional activator GCN4.



(b) Collagen 3-helix: synthetic model peptide containing region of human type III collagen.



(c) Zinc finger: synthetic peptide corresponding to the 31st zinc finger from Xenopus protein Xfin.

(d) λ-cro repressor protein (from bacteriophage λ), with helix-turn-helix motif (residues 16-35) shown in red.

Figure 1.9: Common peptide motifs. Figures made with PyMol [27] using PDB files 2ZTA [93], 1BKV [94], 1ZNF [95] and 6CRO [96], respectively.

needed for such deformation of the helices was expected to be low and hence plausible.

Peptides that form coiled coils have a characteristic sequence, known as the heptad repeat [97; 99], of HPPHPPP, where H represents a hydrophobic residue and P a polar residue. The residues are conventionally labelled *abcdefg* (and hence *a* and *d* are usually hydrophobic). The *e* and *g* residues are often charged, and can form salt bridges between the helices. The coiled-coil structure therefore usually has an inner seam of hydrophobic residues forming a hydrophobic 'core', flanked by *e* and *g* salt bridges, with polar residues on the outer surface. Coiled coils with all of the helices' N-termini at the same end are known as parallel; those with some helices running in the opposite direction are antiparallel [99]. The individual peptides in a coiled coil may or may not be helical in isolation; some coiled coil-forming peptides are only helical when stabilised by the presence of an appropriate partner (forming heteromeric coiled coils), or are self-stabilising, forming homo-oligomeric structures.

Experimental determination of individual structures, searching protein databases for peptides with coiled-coil characteristics, and designs of new peptides based on rules implied by the above studies have drastically increased the understanding of the sequences leading to the formation of different coiled-coil oligomers.

The Protein Data Bank (PDB) [100], established in 1971, contains structure files (generally determined by NMR or X-ray diffraction) for many protein and DNA/RNA-based molecules, and is freely available, making it a great resource for large scale searches/analyses.

Several algorithmic programs have been developed to extract coiled coils from protein sequence databases using different search criteria. Some look at the occurrence of certain residues in particular heptad positions (COILS [101], PAIRCOIL [102], PAIRCOIL2 [103]), some use Hidden Markov Models (MARCOIL [104], CCHMM [105]), and others use these plus additional data, such as amphiphilicity of sequences (SOSUIcoil [106]), or searching for (using algorithms such as PSIBLAST [107]), and using, sequences which are evolutionarily similar (CCHMM-PROF [108], PCOILS [109]), to identify coiled-coil motifs. There are then programs which try to predict oligomeric state from coiled-coil sequences. SCORER (now upgraded to SCORER 2.0 [110]) scores sequences against amino acid profiles, to differentiate between parallel dimers and trimers, while MultiCoil [111] and SPIRICOIL [112] relate sequences to known structures in order to predict oligomeric state (only between parallel dimers and trimers for MultiCoil). SOCKET takes a different direction to finding coiled coils - it looks for the characteristic knobs-into-holes (KIH)

packing in known structures in order to select coiled coils from the PDB [113].

Natural Coiled-coil Structures

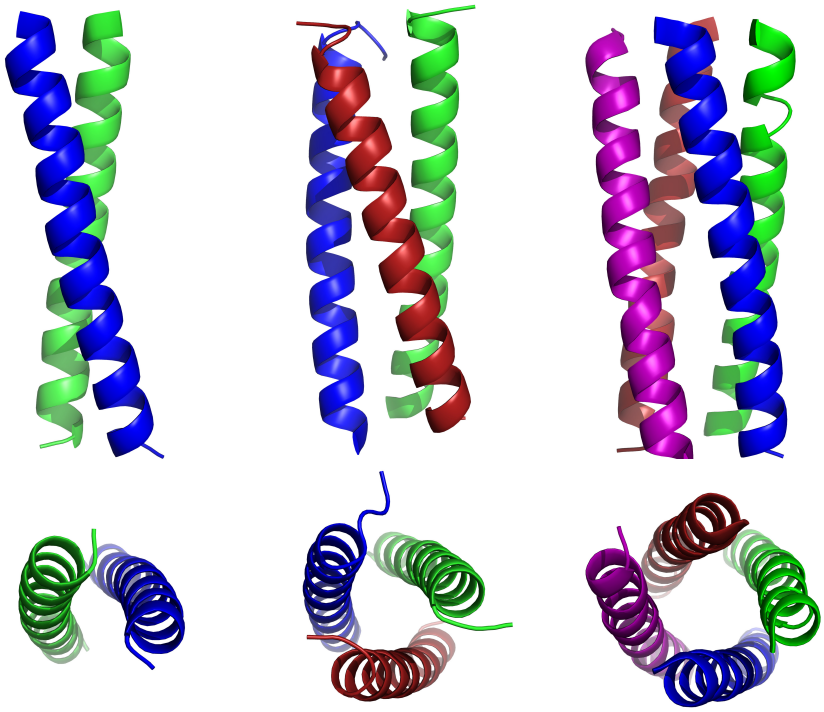
Two, three, four and five-helix classical coiled coils have been found naturally [114]. No natural classical hexamer has been seen, but is feasible due to the existence of a naturally occurring complex coiled coil which includes a six-helix coiled coil [115], and the synthesis of a designed hexamer [116]. Complex coiled coils are structures formed by more than one classical coiled coil (and hence have multiple hydrophobic cores) [114], and though less common, occur in various forms in the PDB, the most common being two- and three-dimer constructs. Moutevelis & Woolfson review the coiled coils seen so far in their 2009 paper [114]. Here we concentrate on the most common structures, the dimer, trimer, and tetramer (Figure 1.10). It is minor differences in sequence that can result in oligomerisation switches between these three, and these tend to be the structures most commonly used in *de novo* design.

Knobs-into-holes Packing

Before discussing the residues which somewhat specify the oligomerisation state of coiled coils, it is useful to understand the means by which the helices pack together. Knobs-into-holes (KIH) packing, coined in 1953 by Crick, is characteristic of coiled coils. Leucine zippers, known as such due to their possessing a leucine repeat every heptad [119], form coiled-coil structures and are found in several transcription factor proteins. It was the X-ray structure of the homodimeric GCN4 leucine zipper published in 1991 which confirmed Crick's KIH packing (Figure 1.12) for a two stranded parallel coiled coil [93].

As previously mentioned, the a and d residues of each helix form the hydrophobic core. Coiled-coil helices are all individually right handed, but supercoil around each other left handedly, reducing the residues per turn (in supercoil space) from the characteristic α-helical 3.6 to 3.5. The heptad repeat therefore covers two turns (Figure 1.11), so each 'layer' of a coiled-coil helix has either an a or d residue close to the inner seam.

It is worth noting at this point that right-handed coiled coils do also exist, but are far less frequently observed. However, a natural example of a tetrameric right-handed coiled coil has been found in the tetrabrachion protein of *Staphylothermus marinus* [120], with an 11-residue repeat, covering 3 turns (3.67 residues/turn), as opposed to the usual 2 turns for left-handed coiled coils. It has leucine and isoleucine residues at a and h positions. A



(a) GCN4-p1 dimer (b) GCN4-II trimer (c) GCN4-LI tetramer

Figure 1.10: Dimeric, trimeric and tetrameric classical coiled coils. The dimer shown is the original GCN4-p1, a leucine zipper from the yeast protein GCN4. It has leucine residues at *d* positions, and 4 out of 5 of its *a* residues are also hydrophobic. Mutants GCN4-II and GCN4-LI both have isoleucine residues at their 1st-4th *d* positions, and at their (2nd-5th) *a* positions, have isoleucine, and leucine, residues respectively. The mutant GCN4-IL (not shown, with isoleucine *a* and leucine *d* residues), appeared to be dimeric in nature, like the wild-type peptide [117]. Figure made using PyMol [27] and PDB entries 2ZTA [93], 1GCM [118] and 1GCL [117].

designed tetramer has also been created [121] using an 11-residue repeat; both of these structures are less supercoiled than equivalent left-handed coiled coils.

In a typical left-handed parallel dimer, the two helices are rotationally offset 180° from each other, and the *a* and *d* residues face inwards towards each other, and in the helical axis dimension, towards the N-terminus (Figure 1.13) [99]. The *g* and *a* residues are approximately in the same layer as each other, while the *e* and *d* residues are in the other. Dimers can therefore, interact with pairwise-complimentary KIH packing [122]: an *a* knob in Helix 1 fits into a hole formed by the d_{-1}, g_{-1}, *a* and *d* residue side chains of Helix 2 (where d_{-1} is the *d* residue of the previous heptad), but this Helix 1 *a* knob also forms one side of a hole for a knob on Helix 2. *d* knobs fit into holes formed by *a*, *d*, *e* and a_{+1} residues of the opposite helix.

In higher order coiled-coil oligomers, the interactions cannot be pairwise; there is instead cyclic complementarity [122]. For a trimer, an *a* knob from Helix 1 fits into a hole in Helix 2, while an *a* knob from Helix 2 fits into a hole on Helix 3, and an *a* knob from Helix 3 fits into a hole in Helix 1. With the same helix labelling, the *d* knobs act in the opposite direction: a Helix 1 *d* knob fits into a Helix 3 hole, a Helix 3 *d* knob fits into a Helix 2 hole, and a Helix 2 *d* knob fits into a Helix 1 hole. This opposing directionality applies to tetramers as well as trimers.

Peripheral KIH packing is seen in tetramers (and somewhat in trimers, for residues with sufficiently long side chains), where *e* and *g* residues can also be knobs which pack into holes formed in neighbouring helices (formed by *c*,*d* and *a*,*b* residues respectively). There are essentially two heptad repeats superimposed on each other, offset by one residue. For a given helix, the *e* knobs interact with the same neighbouring helix as *a* knobs, and *g* knobs interact with the same helix as *d* knobs.

This all means that there are separate *a* and *d* layers in parallel coiled coils (Figure 1.12). Dimers and tetramers have alternating layers of parallel and perpendicular packing. However, it is the *a* residues which have parallel packing in dimers, and perpendicular packing in *d* layers, while tetramers have the inverse. Trimers have acute packing angles for all, intermediate between parallel and perpendicular packing angles.

Antiparallel oligomers have mixed *a* and *d* layers and hence have different KIH packing, and participating residues.

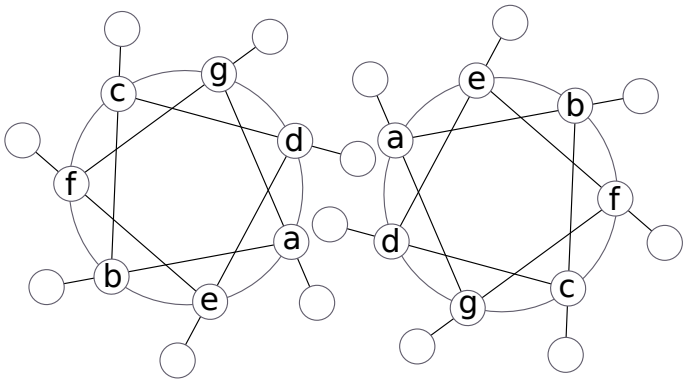
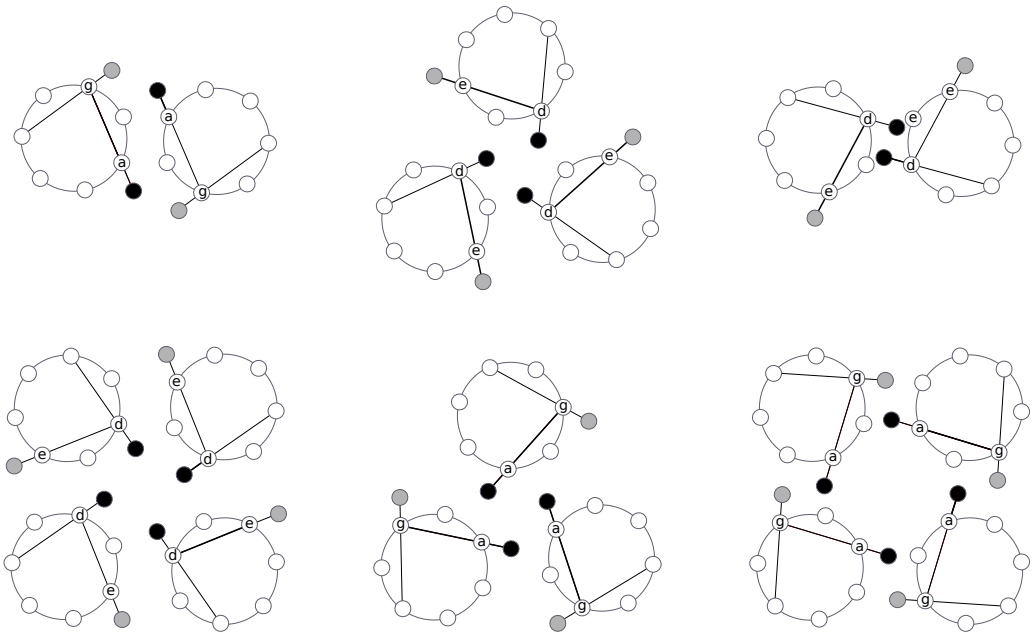


Figure 1.11: Coiled coils have seven residues per two turns, and hence the approximate directions of the *a*-*g* sidechains (white circles) can be displayed on a 'helical wheel'.

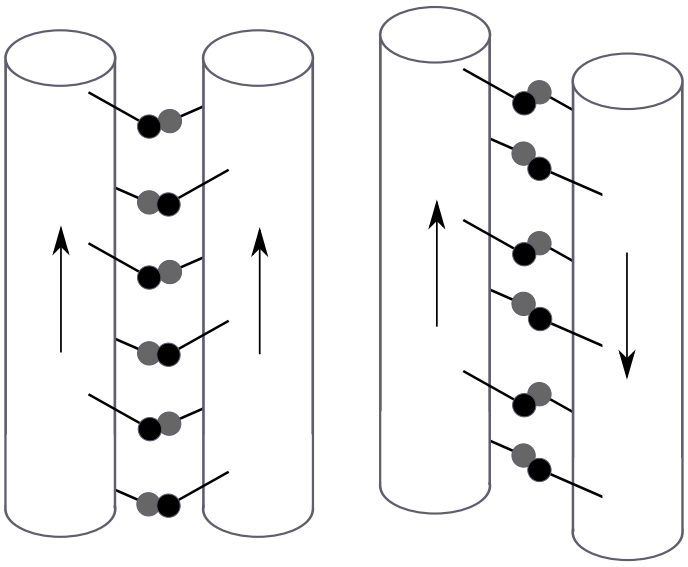


(a) There is parallel packing in the dimer *a* layers and tetramer *d* layers.

(b) Parallel trimers have acute packing, an orientation between parallel and perpendicular, in both the *a* and *d* layers.

(c) There is perpendicular packing for dimer *d* layers, and tetramer *a* layers.

Figure 1.12: The knobs-into-holes packing [97] of parallel coiled coils. The grey circles are the side chains in the shown layers contributing to a hole, while the black circles are both knobs, and part of a hole for another knob. Parallel and perpendicular packing is determined by the approximate angle between a knob's C_α (white circle)-C_β(black circle) vector, and the C_α-C_α vector between the knob's hole residues in that layer (in the (a) dimer and (c) tetramer, the vector between the white 'g' and the white 'a' circles).



(a) Parallel Coiled Coil (b) Antiparallel Coiled Coil

Figure 1.13: Parallel and antiparallel structures. Arrows indicate the direction from N- to C-termini. All side chains point towards the N-terminus, and hence the packing is different between parallel and antiparallel structures. While parallel oligomers have separate a and d layers (discussed in Figure 1.12), antiparallel oligomers have mixed a and d layers [99].

Dimers, Trimers and Tetramers - Specifying Oligomeric State

In addition to confirming KIH packing, the GCN4 leucine zipper ('GCN4-p1'), a 45 Å long, 30 Å wide dimer [93], and subsequent mutations of this 33 amino acid sequence, revealed several of the now famous 'rules' for coiled-coil formation.

A central asparagine residue at an a position helps specify for dimericity

GCN4-p1 has a clear leucine repeat at its d positions (the definition of a leucine zipper), but not all hydrophobic a residues. Four out of five are hydrophobic (three valine, one methionine), but the central a position is occupied by a polar asparagine (N) residue. A hydrogen bond forms between the asparagine residue in each of GCN4-p1's helices [93]. This occurrence of non-hydrophobic a residues was also found in several other leucine zippers [123] [119], some having more than one non-hydrophobic a residue. This property led O'Shea et al. [93] to suggest that they must perform a function; it was proposed that in causing instability, they allowed specificity to be imposed. A mutant with valine replacing the asparagine residue supported this; it had a higher melting temperature than GCN4-p1, indicating higher stability, and did not solely form dimeric species.

Isoleucine prefers parallel packing

Harbury and coworkers [117] mutated the (2nd-5th) *a* and (1st-4th) *d* residues of GCN4-p1 to valine, isoleucine or leucine (for example, GCN4-IL had isoleucine at *a* positions and leucine at *d* positions). The valine containing mutants (-VI, -VL and -LV), and GCN4-LL did not form one specific oligomeric structure. Mutant GCN4-IL formed a parallel dimer, GCN4-II a parallel trimer and GCN4-LI a parallel tetramer. The swapping of leucine and isoleucine between dimeric GCN4-IL and tetrameric GCN4-LI corresponded to a swap in the packing of the *a* and *d* layers (Figure 1.12); isoleucine prefers parallel packing. This packing preference was proposed as the reason for GCN4-II's trimeric structure; as a dimer or tetramer, half of the *a* and *d* isoleucines would have to pack in a perpendicular fashion - acute packing (a compromise) is more preferable [118].

Designed peptides containing one asparagine per peptide in an otherwise hydrophobic (leucine) core formed a heterodimeric parallel coiled coil [124], and without the asparagine a heterotetramer with more than one arrangement resulted [125]. The parallel nature of the asparagine-containing coiled coils is likely due to an antiparallel orientation forcing polar asparagine residues to pair with hydrophobic leucine residues [93], which is energetically unfavourable compared to asparagine-asparagine interactions.

Reasons for dimer formation appear to be amino acid specific; it is not an overall polar group characteristic

Further studies swapping Asn-16 for aminobutyric acid [126] and glutamine [127] resulted in dimer-trimer mixtures, while lysine and norleucine [127] formed dimers, like asparagine (it had earlier been found that glutamine was a frequent occupier of *a* trimer positions, and lysine of *a* dimer positions [128]). However, the norleucine side chains buried into the core while the lysine side chains pointed outwards towards the solvent. A trimeric structure would allow less solvation of lysine, so it appears that maximising solvation favours dimerisation in lysine mutants, unlike the apparent hydrogen bonding driver for the wild-type asparagine coiled coil. The glutamine mutants were again different; in both dimer and trimer form, the glutamine mutants contained one and two water molecules respectively in their coiled-coil cores, but different interactions occurred in the two oligomers. The reasoning behind the dimeric preference of certain residues may even vary for a given residue between different peptide sequences. An intended homodimer, CC-Di, by the Woolfson group [129], needed its initial 'a=I,d=L' sequence to be changed to include a central *a* asparagine to form a dimer instead of a trimer (which it initially formed, in

both crystal and solution forms), but it was not evident that the dimeric favourability of this Asn-containing peptide was due to hydrogen bonding. It was proposed that perhaps the burying of the Asn in a dimer was less so than in a trimer, and hence less energetically disfavoured. Their subsequent synthesis [129] of the design inspiration, GCN4-IL, indicated that it might in fact form both dimers and trimers in solution as opposed to specified dimericity, not disproved by the inability to solve its crystal structure (both by the Woolfson group, and its absence in the earlier work). Alongside CC-Di, the Woolfson group designed CC-Tri and CC-Tet, with sequences identical to CC-Di's original sequence (no asparagine), but with II and LI at *a* and *d* positions; these formed homotrimers and homotetramers respectively, as per the Harbury rules [129].

Other Considerations - Salt Bridges and Steric Constraints

In addition to *a* and *d* interactions, there are also salt bridges and steric considerations which contribute to structure determination. Many coiled coils have charged *e* and *g* residues, and their importance appears to be varied.

In O'Shea, Lumb and Kim's heterodimer design [124], the only difference between the two peptide sequences is that one has positive *e* and *g* residues, and one had negative ones - the *e* and *g* positions were used to specify heteromericity. On the other hand, the designed Coil-Ser [130], intended to be a parallel homodimer [80], in fact proved to be an antiparallel trimer. This was the most stable configuration in terms of hydrophobic potential energy, but had several repulsive interactions between *e* and *g* charges [130] designed to encourage dimer formation (but failing to do so).

e and *g* salt bridges can occur in all oligomers, and although less common, other salt bridges can occur. The GCN4-II trimer [118] had more *e*-*g* bridges than the -IL dimer and -LI tetramer, but they were present in all three. Several *g*-*b* and *c*-*e* bridges were also seen in the tetramer.

Coil-Ser also demonstrates the effect of steric constraints on configurations. The peptide has all *a* and *d* leucine residues with the exception of one tryptophan (for spectroscopy). The layers of hydrophobic residues with two tryptophan and one leucine residue were forced to have one tryptophan facing outwards towards solvent, due to steric constraints [130]. The antiparallel nature of the trimer avoids a steric clash of having three Trp residues in a single layer, which would occur in a parallel trimer [130; 99]. Support for the steric clash discouraging a parallel Coil-Ser is given by the formation of a parallel trimer

[131] by a Coil-Ser mutant with valine *a* residues.

1.4.4 Designing Coiled Coils

In order to use coiled coils as (parts of) components for synthetic constructs, the coiled coils' self assembly needs to drive the components to combine in an ordered, specific, reproducible way.

Channon et. al [132] describe the idea of not just designing a single sequence to form a given structure, but to design what they term 'tectons', a set of sequences which form their own structures and connect together through self assembly to form higher order structures.

As has been discussed, there are some 'rules' for coiled-coil design, but they are more like guidelines; there are exceptions. Some level of helicity is frequently imparted on peptides as a result of designing coiled coils, but it is not necessary for the peptides to be helical in isolation; they may still be highly folded and helical when mixed with an appropriate partner to form a coiled coil. Choosing residues which allow helicity but don't necessarily encourage it nor prevent it may be preferable, as a clear structure change is then visible upon partner interaction.

The frequently seen '*a*=isoleucine, *d*=leucine' [118] with one centrally placed *a* asparagine residue is a popular starting point in dimer design, [93; 117; 128], plus complementary *e* and *g* residues to help stabilise the structure. Glutamine, lysine and glutamic acid are popular *e* and *g* choices. Coiled coils rarely contain proline or glycine as they are 'alpha helical breakers' [133].

Most coiled coils in nature are 'blunt ended' - the peptide chains interact with each other fully. However, it is possible for peptides to interact in a staggered manner, so that one chain interacts with half of two other chains (Figure 1.14). Arranging charged residues so that the first heptads of one chain are conducive to salt bridge formation with the end heptads of another chain in a parallel dimer is one way of achieving this [134]. These 'sticky ended' peptides [135] are particularly useful when forming extended structures such as fibres (Section 1.4.1).

Designing Orthogonal Sequences

Bromley et al. [135] describe a means of designing sets of coiled-coil parallel dimers which are orthogonal, and hence could be used to bring together different components in a

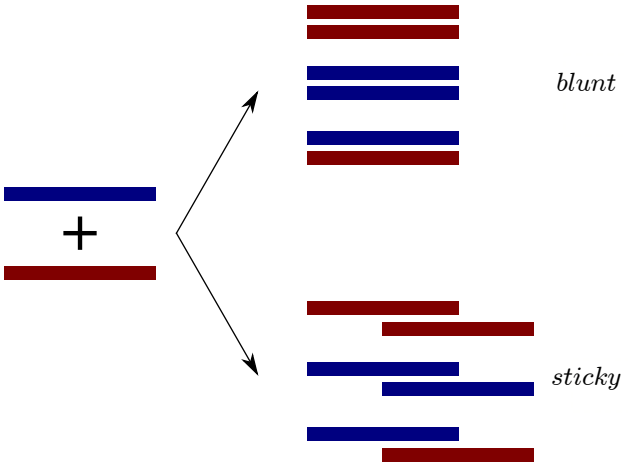


Figure 1.14: The interactions between two peptides can result in many configurations. For a dimeric structure there are several possibilities: homomeric or heteromeric species, staggered ('sticky') or unstaggered ('blunt') coiled coils (as shown here). Each of these six configurations shown can be either parallel or antiparallel.

system into a set configuration, reproducibly, over multiple copies. Figure 1.15 describes their method. The idea is to calculate the interaction energies of all possible sequences in all configurations (Figure 1.14), and maximise the gap in energy between desirable structures and all of the alternatives. From the pairs found using the process shown in Figure 1.15, preference for the final set was given to sequences which also fulfilled the following criteria:

1. 'E at *g*, K at *e*' sequences (as seen in bZIP proteins [136; 137])
2. Positive sequences (for solubility, and hence purification)
3. Sequences that when linked to each other are still orthogonal

The *b*, *c* and *f* positions were then filled to include one tyrosine per sequence, and alanine and glutamine residues, to give sequences of six unique masses [135]. This technique is used in the design of the Tumbleweed peptides.

1.4.5 Coiled-coil Nanostructures

The Woolfson group used another pair of designed parallel heterodimeric coiled-coil peptides, linked together by $(GN)_X$ residues to form a monomer, to demonstrate the use of steric effects (due to both the linker and helices) to control self assembly of structures [138]. Fibres formed for an X=1 linker, and square and triangular constructs for X=3

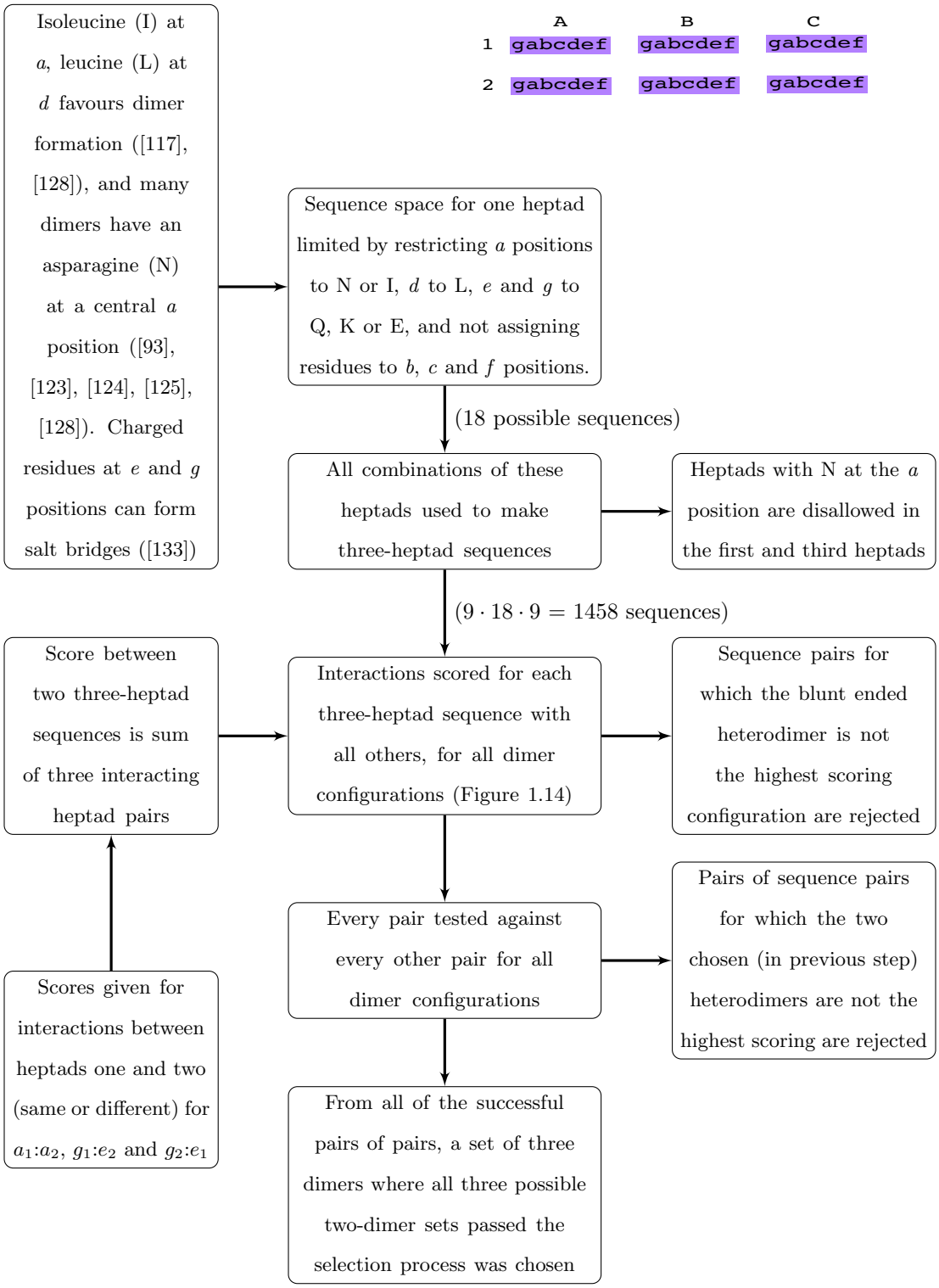


Figure 1.15: The scheme for designing orthogonal coiled-coil dimer sets, as described by Bromley et al. (2009).

and 4 respectively (X=5 formed a dimer-trimer mixture, and X=2, various colloid-like structures). A monomeric structure of the linked pair was prevented by steric constraints due to the peptides' parallel dimer preferentiality, but is likely possible with a sufficient linker length.

Coiled coil interactions have been used to create several more complex structures, and two of the most interesting are spherical, but rather different to each other.

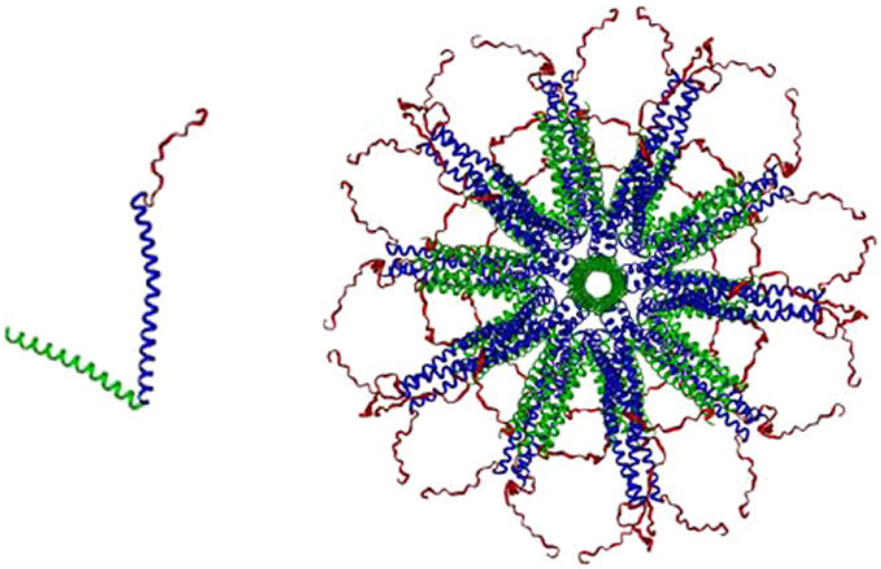
The Woolfson group have used coiled-coil interactions alongside disulphide bonding to produce SAGEs - Self Assembled caGE-like particles (Figure 1.16b) [139]. They use a designed homotrimer and heterodimer. By disulphide bonding homotrimer monomers to either of the heterodimer peptides, they form six-peptide structures driven by the trimer formation, and then, by mixing the two variants of these six-peptide structures, form cages of sizes around 100 nm, driven by the heterodimer formation. The cages are spherical; it is thought that with the disulphide bond acting as a hinge (towards one end of the peptides), repulsion between lysine residues at the other ends of the disulphide bound homotrimer and heterodimer peptides causes them to form a 'V' shape, flexible enough for them to form a sphere with only hexagonal faces.

The Burkhard group have used a *de novo* trimer coiled-coil design alongside a modified coiled-coil domain of COMP (which forms a pentamer) [140] to synthesise a monomer with both domains, separated by a linker. These monomers can be used to form spherical objects [75]. They have then decorated these SAPNs (Self Assembled Peptide/protein Nanoparticles) with epitopes from actin [141], avian influenza [142], malaria (Figure 1.16a) [143], SARS [144] and HIV-related [145] proteins. The particles are of the order of 20 to 40 nm, depending on their decoration, and some SAPN experiments have induced immune responses, a good indicator of their potential for future vaccines [143].

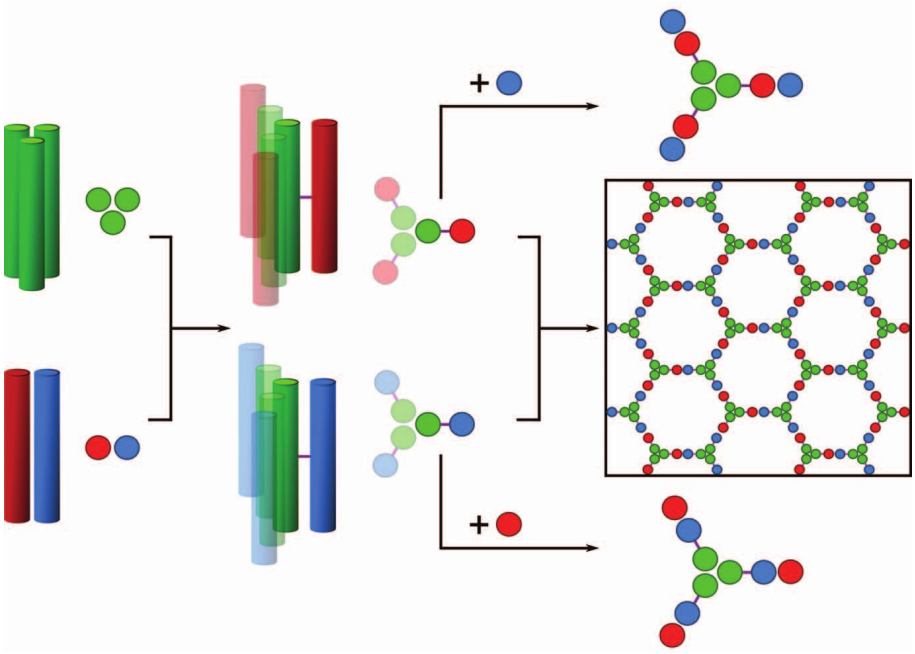
1.5 The Tumbleweed Motor

1.5.1 Concept

The Tumbleweed is a synthetic protein motor design [1]. It comprises of a hub which connects three legs (each with a 'foot'), and this motor, by rotating diffusively, walks along a track, with one or two feet bound at any one time. In the present design, the Tumbleweed comprises of a peptide-based hub (including the legs), with 'feet' made from protein repressors of Escherichia coli, which are DNA-binding, walking along a track made



(a) The malaria-decorated SAPN of the Burkhard group: a single monomer with trimer and pentamer-forming domains, and malaria epitope, and a formed nanoparticle. Image taken from [143].



(b) The formation of SAGEs by the Woolfson group. Image taken from [139].

Figure 1.16: Examples of coiled-coil nanostructures. Copyright notices in Appendix A.

of DNA. In order to control the feet that are bound at any given time/position, the three feet are different repressors, and each binds to a different DNA sequence when a certain different ligand is present. The concept is explained further in Figure 1.17. A possible structure of the Tumbleweed is shown in Figure 1.18d. Directionality will be imparted by having a directional track (A-B-C-A-B-C), and controlling the feet which are bound and unbound through external ligand control. At present there is no power stroke, which, in some natural motors, imparts directionality, only a means of preventing stepping in one direction and allowing it in the other, through regulating binding. The Tumbleweed can diffuse in the wrong direction, but because there are three different binding sites on the track, if the foot behind the one bound sets down, its ligand will be absent and hence it will not bind to the track (in the absence of non-specific binding, discussed in Section 1.5.3).

This PhD project focuses on the hub construction, but it is important to be clear where this fits into the Tumbleweed project as a whole; work has been carried out elsewhere on the DNA track, the repressor feet, and the nanofluidics system, along with simulations that both assisted in the design and continue to gain further results on the system. That work is reviewed here in the rest of this chapter.

1.5.2 Timescales

There are many timescales which need to be considered when designing a motor like the Tumbleweed. At any given time, two ligands will be present in solution. Two given ligands will be present together for a length of time τ_{ligand}, after which one of the two will be replaced by the third ligand of the system. These will then be present in the system together for another τ_{ligand}, after which the ligand which has been in the system for $2\tau_{\text{ligand}}$ is replaced with the ligand which had been absent in the last τ_{ligand}-long 'pulse'. There are therefore three different pulses which occur in sequence, and are then repeated: ligand A with ligand B, B with C, and C with A, each lasting τ_{ligand}.

Consider the situation when there are front and back feet bound to the track, and the third foot is unbound. When the ligands are changed, the back foot (called the back foot for this whole example), whose ligand is no longer present, needs to detect this, react, and detach from the DNA track. This depends on the rate at which the ligand and repressor associate/dissociate ($k_{\text{lig-rep}}$), and the rate at which the repressor unbinds the track once it is no longer bound to a ligand ($k_{\text{off}}^{\text{-lig}}$). The free foot also then needs to find its binding

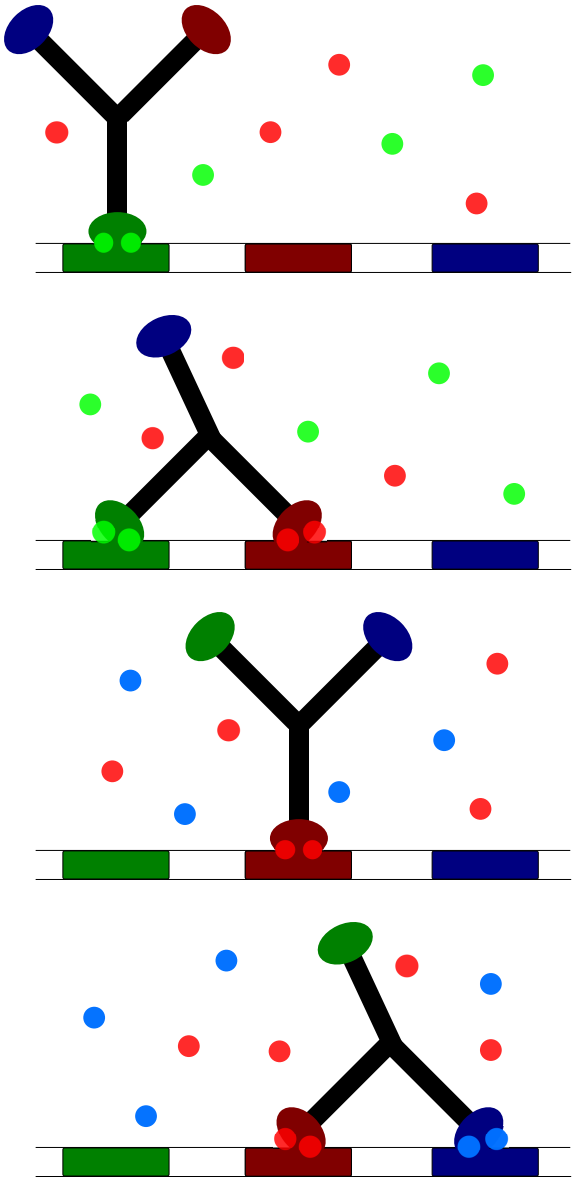


Figure 1.17: The Tumbleweed and its proposed method of motion. The DNA track has alternating patches to which the three different ligand-bound feet can bind. Changing the solution (ligands) present will cause the feet to attach and detach in turn. In Part 1, ligand a has been present for a time, and hence foot A has been able to bind to binding site (BS) A. Ligand b has been added, and so B can then also bind, to BS B (Part 2). Ligand a is then removed, and ligand c added. A lifts up (Part 3), and the motor can diffusively rotate so that C can reach BS C, and bind due to ligand c's presence (Part 4). This process would continue with the removal of b, addition of a, lifting of B and binding of A, followed by the removal of c and readdition of b, plus lifting of C, to return us to the situation in Part 1. Cycle repetition would allow the motor to make multiple rotations/steps.

site and bind (which takes, on average, a characteristic diffusion time τ_{diff}), before the front foot unbinds and lifts. The front foot can unbind due to the next ligand change (τ_{ligand} after the previous ligand change) removing its ligand from solution, or because it naturally dissociates even with its ligand still present (τ_{bound} after binding, on average). If the free foot does not bind before the front foot lifts, the motor will fall off the track, as no feet will be bound. It is therefore important that the combined processes of the back and free feet are quicker than both τ_{ligand} and τ_{bound}. The ligands need to be changed more quickly than the average motor foot naturally dissociates, or there will be periods of the cycle when the back foot naturally dissociates, but the free foot stays unbound as its ligand is still not present, leaving just the front foot bound. This would lead to a greater part of the stepping cycle with only one foot bound, and the shorter the characteristic binding time τ_{bound}, the more likely it is that the single bound foot will dissociate and the motor will be lost from the track. Therefore, a τ_{bound} significantly longer than τ_{ligand} is desirable.

These interlinking timescales are more clearly described by an inequality equation:

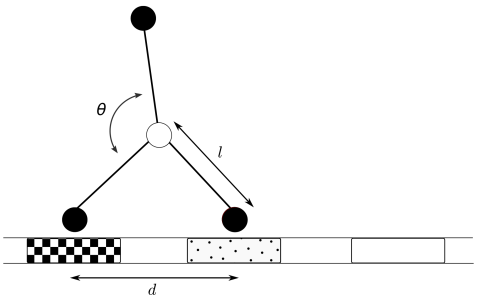
$$\tau_{bound} \gg \tau_{ligand} > \frac{1}{k_{off}^{-lig}} + \tau_{diff} + \tau_{lig-rep} \tag{1.1}$$

where

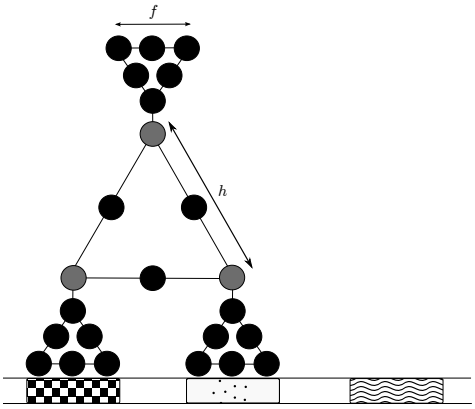
$$\tau_{bound} = \frac{1}{k_{off}^{+lig}} \tag{1.2}$$

$\tau_{\text{lig-rep}}$ is thought to be on the order of picoseconds for the Tumbleweed repressors [146], much quicker than the other timescales. τ_{bound} is thought to be of the order of 100s [1; 146], and $\frac{1}{k_{\text{off}}^{-\text{lig}}}$ at least one hundred times faster (1s) [1; 146], or even 1 ms according to some experimental data [146]. τ_{ligand} (discussed in Section 1.5.4) is on the second timescale [147], and τ_{diff} of the order of $100\mu s$ [1; 146; 148].

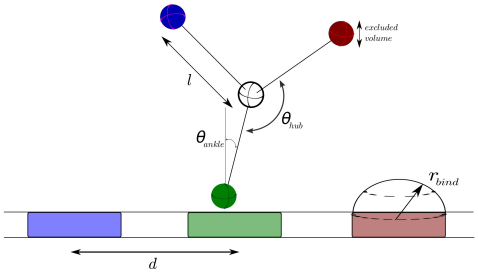
Non-specific binding is seen in natural motors, and is thought to help motors find their next binding site, as they clearly do not search all of the possible space. It may therefore be an advantageous property, but if the non-specific binding is too strong, a foot may not release when its ligand is removed, a problem for a processive motor design. The strength of non-specific binding can therefore affect both the rate at which feet detach from DNA, and find their new binding sites.



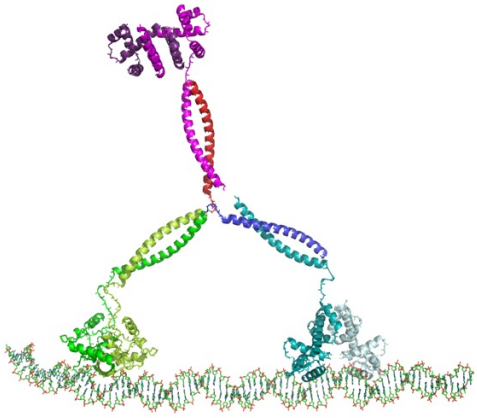
(a) The 2D Y model.



(b) The 3D triangular model.



(c) 3D coarse-grained model for Langevin simulations.



(d) A likely structure of the Tumbleweed, from feet crystal structures and predicted hub structure.

Figure 1.18: The many faces of the Tumbleweed. The 2D model has one single joint, while the 3D triangular model has one flexible joint (shown in grey) for each foot [1]. The 3D Langevin model has each foot represented by a sphere of size and drag coefficient to approximate protein counterparts, and the hub has three legs joined in the centre with another sphere. The leg lengths (r_{ik}) were maintained by harmonic potentials, and the sphere sizes/excluded volumes by repulsive L-J potentials. It has both a flexible hub joint and leg-to-foot (ankle) joints.

1.5.3 Simulations - Non-specific binding, hub flexibility and timescales

Various models of the Tumbleweed concept have been produced to examine the importance of steric and separation constraints, flexibility of joints, and timescales, both as part of the initial design stages [1], and ongoing work on the system [146; 148].

The simplest model is the Y motor model (Figure 1.18a) [1]. A 2D simulation was carried out with an overdamped Langevin equation (also used in simulations in Chapter 4.2, where it is discussed further), with three spherical 'feet' connected by stiff axial spring potentials at a single joint of variable flexibility representing the motor, and a line of binding potentials representing the track. Ligands were supplied for τ_{ligand}-long pulses, of (a,b), (b,c) and (c,a) ligand pairs successively (and repetitively), and foot binding was only allowed during the presence of the correct ligand, with immediate unbinding when the ligand was removed. It was found that, for an inflexible joint ($\theta = 120°$), leg length $l = d/2sin\theta$), the joint could make multiple steps. If the leg was made a little shorter, the motor struggled to bind to two sites at one time and was therefore more likely to detach. Allowing thermal fluctuations around a 120° angle resolved this. As DNA is flexible and the binding process will induce changes, some flexibility in the motor design might be needed to tolerate variations in the DNA binding site separations.

A slightly more complex 3D simulation (Figure 1.18b), with a triangular hub, tackled both flexibility and steric effects. The feet were given a triangular shape, with a screened, repulsive Lennard-Jones potential (the screening length approximating the true repressor size), each attached via a rigid joint to hub joints of variable flexibility. To sterically allow sequential site binding, the length of the hub sides (h) needed to be bigger than those of the foot sides (f). Successful stepping was seen for a hub length twice that of the feet with a τ_{diff} of $150\mu s$.

For more complex simulations, other timescales were introduced. In earlier simulations, $\frac{1}{k_{\text{off}}^{\text{-lig}}}$ was taken as simultaneous with the ligand change. It needs to be included if one wants to be realistic and investigate non-specific binding [146; 148].

A study on the effects of timescale relationships and ligand switching efficiency on stepping using a classical master equation (which describes transitions between states), with input of results from previous Molecular and Langevin Dynamics models [1], gave the likely binding state of motors at a given time, from which stepping/lack of stepping could be inferred [146]. Transitions which gained or lost one foot binding were allowed; a two-foot-bound motor could not detach, nor a non-bound motor bind multiple feet, in one

transition. Matrices of rates for ligand binding transitions at a constant DNA binding configuration and vice versa were built to describe motor state changes. τ_{diff} and τ_{ligand} values of 200 μs and 1s were used, respectively, along with a τ_{bound} value of 100 s and $\frac{1}{k_{\text{off}}^{\text{-lig}}}$ of 1 ms. For a fixed τ_{ligand} and varying τ_{bound} (0.1-100 s) and τ_{diff} (2, 20, 200, and 2000 μs), it was found that, as one might expect, motors were more likely to detach from the track (within the first 30 steps) as τ_{bound} decreased (as it was increasingly likely for a foot to detach undesirably), and as τ_{diff} got longer (due to increased binding site searching time, with only one foot bound). For microsecond τ_{diff} timescales, the motors were still quite successful even as τ_{bound} decreased, with less than 10% of motors lost after 30 steps for $\tau_{\text{bound}} \geq \tau_{\text{ligand}}$. However, for a 20 ms τ_{diff}, and $\tau_{\text{bound}} = \tau_{\text{ligand}} = 1s$, this loss increased to 20%, and rose dramatically to 90%-plus for $\frac{\tau_{\text{ligand}}}{\tau_{\text{bound}}} = 5$ (compared to $\leq 40\%$ loss for all of the other τ_{diff} values investigated). τ_{diff} is evidently important in determining continued attachment success even when three orders of magnitude shorter than τ_{ligand} and τ_{bound}. Another realistic issue is that there will be a finite changeover time between one pulse and another, when all three ligands will be present: a continuous change from 100% pulse 1 to 100% pulse 2 will occur. Kuwada et al. [146] define a time, τ_2, over which this change takes place (100% pulse 1 to 100% pulse 2), so that:

$$\tau_{ligand} = \tau_1 + \tau_2 \tag{1.3}$$

where τ_1 is the time at which one pulse is at 100%. To reduce computation time, they modelled τ_2 as a step of 50:50 pulse 1-pulse 2 mixture between τ_1 of (a,b) and τ_1 of (b,c) rather than continuous gradients. For fixed $\tau_{\text{ligand}} = 1$ s, the probability of foot A and foot B together being bound during τ_1 of a (b,c) pulse is zero, but is increasingly likely as τ_2 progresses; the longer τ_2 is, the higher the probability (0.1 for $\tau_2 = 0.9\tau_{\text{ligand}}$, versus $\leq$ 0.05 for $0.3\tau_{\text{ligand}}$). The misstep probability increases with τ_2 too; a longer τ_2 corresponds to a shorter τ_1, so less of the AB-bound motors will have stepped to BC (either by lifting A during (b,c) but not binding C before A can rebind, or by stepping to BC and back to AB again during τ_2). For fixed $\tau_{\text{ligand}} = 0.1$ s, the probability of AB still goes from one to zero as a (a,b) pulse moves to a (b, c) pulse if $\tau_1 = \tau_2$. If τ_2 is longer than τ_1, neither a zero nor one probability occurs and missteps occur, because some motors don't lift a foot in τ_1. In the 1 s τ_{ligand} simulation, there are always missteps due to τ_{ligand} being comparable to τ_{bound}. The difference is that in $\tau_{\text{ligand}} = 0.1$ s, although there is basically no misstepping for $\tau_2 \leq \tau_1$, in the region they do occur, the misstep fraction becomes much higher than the $\tau_{\text{ligand}} = 1$ s fraction ever does. $\tau_{\text{ligand}} = 0.1$ s goes from 0 to over 25% misstepping

between $\tau_2 = \tau_1$ and $\tau_2 = \tau_{\text{ligand}}$, while for $\tau_{\text{ligand}} = 1$ s, for the whole range ($\tau_2 = 0$ to τ_{ligand}), misstepping is never higher than 5%.

Even though speed of the motor is technically dependent on τ_{ligand}, if τ_{ligand} is too short, it may not speed up the motor, as payment will be made in the form of misstepping.

A later study by Kuwada et al. [148] looked at tuning the motor to optimise τ_{diff}, by adjusting flexibility and non-specific binding (Figure 1.18c). They concluded that there was no advantage to flexible legs, and that flexible ankle and hub joints, and some non-specific binding, minimised τ_{diff}. They found that negative effects on τ_{diff} due to high non-specific binding could be alleviated by some ankle rigidity, or, in the case of a rigid ankle, some non-specific binding was helpful. As flexibility of the ankle region is thought to be harder to both predict, and modify, it was suggested that adjusting ionic strength to change non-specific binding levels might be the most feasible means of optimising a given motor τ_{ligand}.

1.5.4 System Components

Track

DNA was chosen for the track material due to its mechanical and self assembling properties, and its ability to code for particular binding reactions. The DNA track needs to have [149]:

1. repeating sections of the binding recognition sequences for the three repressor feet, spaced so that they a) correspond to the step size of the motor (11 nm), and b) all occur on the same side of the helical DNA molecule.
2. a means of being anchored and held on a surface.
3. sufficient length to study motor dynamics.
4. a label, so that the motors' progress can be tracked, and separated from any drift of the whole system (the hub should also have a label for tracking).

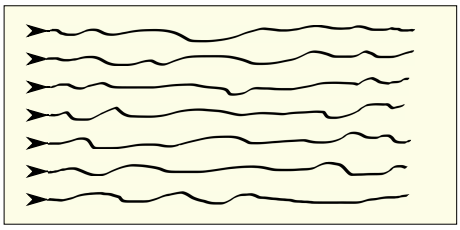
A DNA sequence of the form L$(ABC)_N$R (the 'cassette') was created, where A, B, and C are the chosen repressors' recognition sequences (plus some bases to create appropriate spacing between them), which are repeated N times, and L and R are primers which include sequences for restriction enzymes (to cut the DNA, allowing ligation and doubling of the cassette). A genetic algorithm was used to design a suitable sequence of this form, with desirable thermodynamic properties (a melting temperature of 56°C was aimed for), and

no unintended additional recognition or restriction sites. The repetition in the sequence was also minimised, as bacteria tend to remove repeat sequences. The choice of an E. coli strain ($DH5\alpha$) more tolerant of repeats than other strains was also made to tackle this potential issue. A region of the DNA was labelled with fluorophore. Once the initial cassette had been produced, L$(ABC)_1$R was used to produce tracks of higher N values; L$(ABC)_1$R was put into a plasmid, and using restriction enzymes and ligation, another copy of the ABC sequence was inserted, increasing N to two. The plasmids were then amplified in the E. coli, and this process was repeated to double N multiple times.
A track of L$(ABC)_8$R has been successfully produced, and extended to a sufficient length for a DNA curtain experiment (Figure 1.19a), of 14 μm, by the addition of λ DNA (which has no binding sites for the repressors) [149].

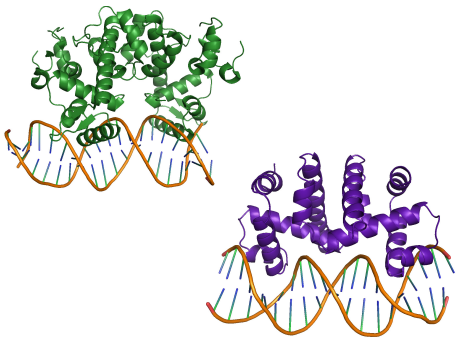
Microfluidics

The microfluidic system for ligand delivery [147] needs to be able to provide the (a,b), (b,c) and (c,a) pairs of ligands in order, repetitively [1], to motors on a surface bound track [149]. It is required that the switching time (similar to τ_2 in Section 1.5.3, but defined as the time between switching from 90% of one solution to 90% of another, as opposed to 100% to 100%) is much less than the ligand pulse time, and for fluid velocity to be low enough so as to not affect the motors' motion negatively.
The resultant device is shown in Figure 1.19d, a PDMS (polymer) device, made from a lithographically patterned mould of SU8 2050 on a silicon wafer, bonded to a glass cover slip for TIRF (total internal reflection fluorescence) imaging, and with silicone tubes attached to holes into inlets/outlets, connecting the system to solution vials. The device is operated so that all three fluids flow continuously, around their own U-shaped channels, and by increasing the pressure on a given outlet (relative to the others), control of which solution is carried through the main channel to the Tumbleweed system is allowed. The system is capable of cycling the solutions in both I-II-III and III-II-I orders, and the fluid switches are well defined even near the surface boundary; the boundary does not hamper the switching times. The interfaces between the different solutions smeared out as they traversed the main channel, but the switching time was as low as 0.2 s close to the surface at the inlet end of the main channel; however, at the pressures used, for the fluid to reach the main channel away from the inlet branches before switching, the system switching time is 1 s. Tests of force on a quantum dot for a 100 $\mu m s^{-1}$ flow found drag to be around

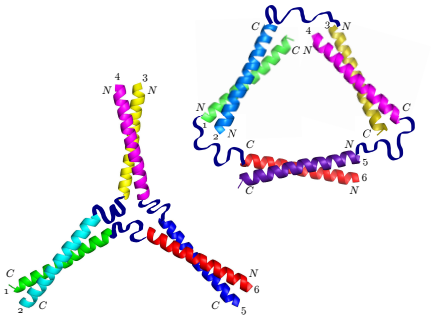
0.02 pN, lower than the 0.5-1 pN it is expected that the Tumbleweed can step against [148], so the system-induced drag should not overwhelm the Tumbleweed's motion.



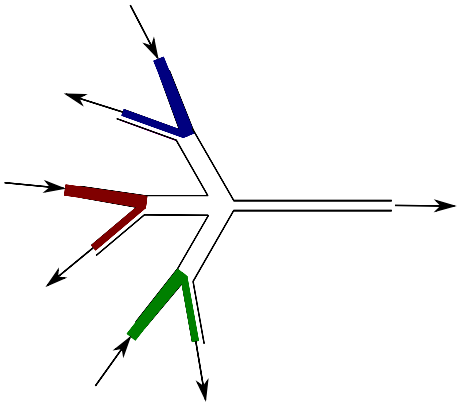
(a) DNA Track [149]: several strands of 14 μm DNA (to form a DNA curtain) have been attached to a surface using biotin and streptavidin.



(b) Repressors TrpR (purple) and DtxR (green), rendered in PyMol [27] from PDB structures 1TRO [150] and 1C0W [151].



(c) The 'Y' and triangle configurations of the Tumbleweed hub (taken from [1]). Both have three dimeric coiled coils; the 'Y' configuration has two parallel coiled coils and one antiparallel, while the triangle configuration has three parallel coiled coils. The thin lines linking the coiled coils represent the bonds which are currently covalent disulphide bonds between cysteine side chain thiol groups.



(d) Microfluidic Chip [147]: the three inlets (left) have the three ligand solutions (shown in red, blue and green) flowing in and out of them. An increase in the pressure of one allows that solution to fill the main area and flow through the outlet to the rest of the system.

Figure 1.19: Progress on the components of the Tumbleweed outside/prior to this project.

Feet

The feet must have DNA binding properties that are highly influenced by the presence of specific ligands [1]. All three feet need to bind to a different DNA sequence in response to binding to a different ligand, to ensure orthogonality. For molecular biology purposes, monomers or dimers were required, and searching the PDB [100], three E. coli repressors were originally selected. Since publication of the Tumbleweed paper [1], the chosen repressors have been revised, due to issues such as uncontrollable non-specific binding. At present, TrpR remains, while DtxR and another as yet undetermined repressor have replaced PurR and MetJ (Figure 1.19b).

TrpR is a homodimer formed of two 107-amino acid subunits [86], and represses L-Tryptophan synthesis when the Trp concentration is high enough that it binds to TrpR, activating it, through a conformational change [86; 79] (seen in crystal structures of TrpR with and without Trp [152]) and allowing the dimer to bind to the major groove of DNA. Trp increases the TrpR dimer binding by a factor of a thousand [152].

DtxR is a homodimer [153; 154], composed of two 226-residue chains [155], and is activated *in vivo* by Fe^{2+} [156], but has been observed *in vitro* to be activatable by other divalent transition metal ions such as cadmium, cobalt (as is the case in Figure 1.19b), manganese, nickel and zinc [157]. Its DNA-binding capabilities appear to be quite specific to its recognition sequences, and strongly dependent on metal ion concentration [158; 159]. The DNA binding domain contains a helix-turn-helix motif, and binds to the major groove of DNA [160].

Hub

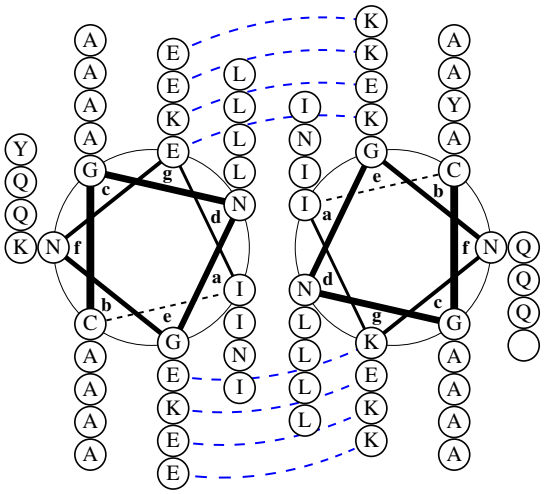
The hub design is for three dimeric coiled coils in two possible configurations to bind with each other and the repressors to form a motor capable of walking along a track. There is a triangular design, and a Y-shaped form (Figure 1.19c). The peptides are designed to form dimers; more specifically, they are six peptides designed to interact specifically with a certain partner, with far more efficiency than any of the other present peptides, to form three different dimers (using the technique discussed in Section 1.4.4). The hub peptide sequences are shown in Table 1.1. Three peptides (1, 4 and 6) are common to both hubs. The other three differ. In the Y hub, peptides 1 and 2, and 3 and 4, form parallel dimeric coiled coils, while 5 and 6 form an antiparallel dimer. All three pairs (1 and β, γ and 4, and ϵ and 6) in the triangular hub form parallel dimers. At one end of each peptide there

Table 1.1: Tumbleweed Peptide Sequences (all written from N to C-termini, which are labelled in Figure 1.19c). Further details in Appendix C.

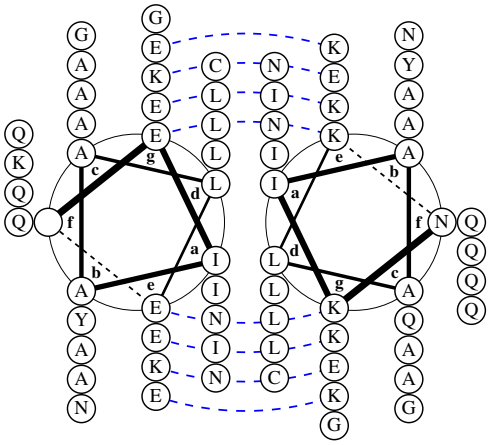
Peptide	**Sequence**
1	CGNGNEIAALEKKIAALKQENAALEQEIAALEY
2	CGNGNKIAALKQEIYALEQKNAALKQKIAALK
β	KIAALKQEIYALEQKNAALKQKIAALKQGNNGC
3	EIAALEQEIYALEQKNAALKKEIAALEQGNNGC
γ	CGNGNEIAALEQEIYALEQKNAALKKEIAALE
4	NKIAALKQKIAQLKQENAALEQKIYALKQGNNGC
5	CGNGNKIKALKQEIAALEYEINALEQ
ϵ	CGNGNEIAALEYEINALEQKIAALK
6	NKIAALKYKIAALKQEIAALEQGNNGC

is a linking region made up of one cysteine, and asparagine and glycine residues (similar to the peptides in [138]). Prior to mixing all six peptides of a hub, the peptides 6 and 1, 3 and 2, and 4 and 5 (or β to γ, 4 to ϵ and 6 to 1) will be disulphide bonded together via their cysteine residues, so that once the peptides are mixed and coiled coils form, the whole hub structure will be complete (Figure 1.19c).

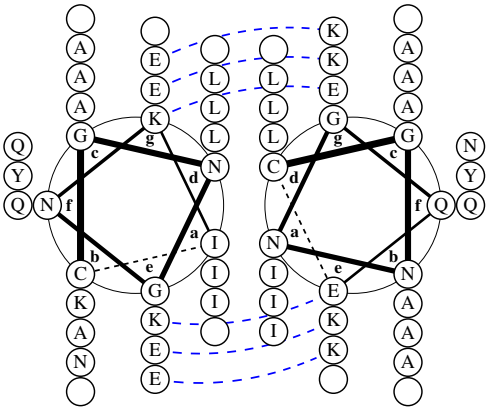
As is clear from the sequences (Table 1.1), β and γ are analogues of peptides 2 and 3, differing only in the positions of their linking regions. This is so that in the Y hub they can form disulphide bonded 3 (N to C) - 2 (N to C), while in the triangular hub they form 2 (N to C) - 3 (N to C). Peptide ϵ varies from peptide 5 more greatly, as ϵ,6 is parallel, and 5,6 is an antiparallel dimer. This project has focussed more heavily on the Y hub, and hence the triangular hub is only discussed here for completeness, as it was also involved in some of the early simulation results. The Y hub offers greater functionality; it has a central vertex for further modification, and the geometry is more open to adaptations, such as a modification allowing inducible conformational changes (a possible means of introducing a power stroke into the Tumbleweed). The peptide sequences follow several of the rules discussed in Section 1.4.3. Apart from in the linking regions (GNNGC or CGNGN), all of the peptides have a leucine heptad repeat (every d residue is leucine). The first four peptides (and their analogues) have four heptads, with an asparagine a residue near the centre of the helix, and the rest of the a residues are isoleucine. The last two (and ϵ), the



(a) 1,2 (parallel)



(b) 3,4 (parallel)



(c) 5,6 (antiparallel)

Figure 1.20: Helical wheel diagrams (made using DrawCoil [161]) showing the interactions between pairs of peptides in the Y hub of the Tumbleweed. Dashed lines show salt bridges between e and g residues. The blank circles are residues absent from the peptide, but which have a corresponding residue in the partner (for example, peptide 1 has one more residue than peptide 2, ending with an f residue, while peptide 2 ends with an e residue (see Table 1.1).

three-heptad peptides, have all isoleucine a residues. All of the e and g residues are either lysine (positively charged) or glutamic acid (negatively charged), to maximise the number of salt bridges (Figure 1.20). The majority of the b and c positions (38 out of 44) are alanine. There are no proline residues present and glycine is only present in the linking regions of the peptides. Cysteine is solely used at the ends of the peptides for its disulphide bonding property, and there is only one per peptide to constrain the possible disulphide bonds that can form. All of the peptides have a tyrosine residue to assist concentration determination and HPLC purification.

1.5.5 Properties Checklist

To summarise, by comparison with the checklist of required and desirable motor properties:

1. The Tumbleweed has a method of being directional.
2. The Tumbleweed should be processive with its current design.
3. The Tumbleweed's structure should be stable enough that it does not fall apart.
4. The Tumbleweed responds to its environment.

The Tumbleweed is not:

1. autonomous.
2. able to carry a cargo in the current design.

Its efficiency could be increased if it had a power stroke, an obvious future aim.

1.6 Thesis Outline

The keys areas of this thesis are the development of the Tumbleweed hub, and work towards another motor design, the Bar motor.

In Chapter 2, I discuss the experimental techniques used in the later chapters: the basic theory behind the techniques, and how they are used to gain information on our systems. Chapter 3 is the first results chapter, containing the results of the Tumbleweed experiments. The individual hub peptides are characterised, their dimeric interactions with each other, and the bonding of the peptides in order to form a hub are studied. Chapter 4 discusses the Bar motor; the concept, peptide design, and simulations of the system are

detailed. Chapter 5 describes experiments on the first iteration of the Bar motor. The thesis is completed with conclusions, possible future directions, and appendices, including experimental methods. In cases where a person other than the author provided a sample or collected data, these are acknowledged within the chapters.

Chapter 2

Theory of Experimental Techniques

This chapter is intended to describe the principles of the chemical and biophysical techniques used in the later chapters of this thesis, and the information we wish to glean about our systems from each of them, as many of the techniques are common to both systems.

2.1 Production of Purified Peptides

The first step in the experimental work of each of the systems was to synthesise and purify designed peptides. All of the peptides in this PhD thesis were synthesised via Solid Phase Peptide Synthesis (SPPS) and purified using High Performance Liquid Chromatography (HPLC).

2.1.1 Solid Phase Peptide Synthesis (SPPS)

Solid phase peptide synthesis was pioneered by Bruce Merrifield [162]. Previously, peptides had been synthesised in the liquid phase, since the first peptide syntheses of Emil Fischer at the start of the 20th century [163; 164; 165].

In SPPS, peptides are formed on resin beads. We use Rink Amide resin; the resin has an $-NH_2$ group, which the first amino acid of each chain bonds to, via its carboxyl group, in the same condensation reaction that occurs between amino acids to form peptides (Figure 1.7). This means that peptides are built up with the C-terminus attached to the resin. The steps involved in solid phase peptide synthesis are shown in Figure 2.1. The employment of microwave energy to assist in the coupling of the amino acids generally makes

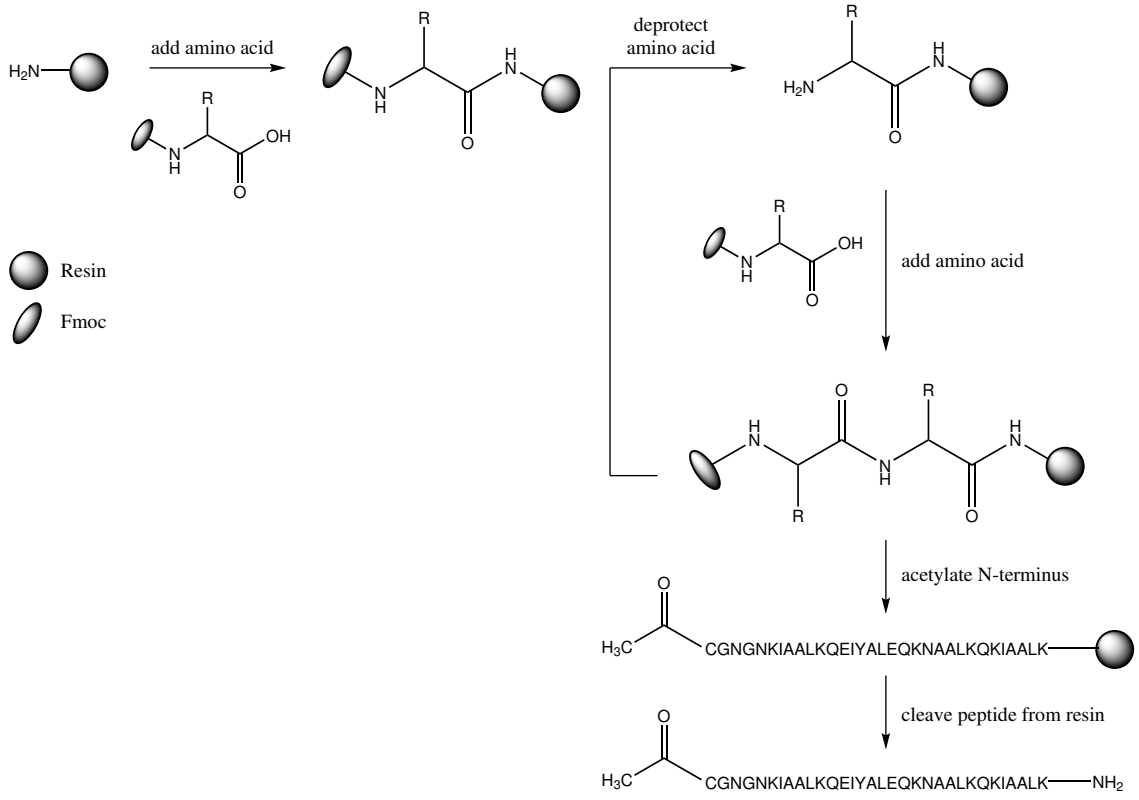


Figure 2.1: The process of solid phase peptide synthesis, using Fmoc protection. The sequence shown here is that of peptide 2, but the sequence is the only difference between the peptide syntheses - the synthesis reactions are the same.

reactions faster, and increases the yield of the correct peptide [166]. This improvement is likely due to the increased temperature, which is known to increase reaction rates in peptide synthesis, but it has also been suggested that the means of interaction between microwaves and molecules helps focus the energy transfer, increasing rates, and that electric field effects on the polar backbone and N-terminal amine help discourage aggregation [167], a cause of decreased yield due to sterically hindering reactions. Microwave energy has been found to increase racemization in cysteine, histidine and aspartic acid residue reactions [167; 168; 169], but this is largely bypassed in this project due to the absence of histidine and aspartic acid, and minimal use of cysteine. It is difficult to estimate the likelihood of a particular synthesis being successful. Particularly by the time lengths of 30 to 40 amino acids are reached, success rates are highly sequence dependent [170]. The particular amino acids used, and their position in a sequence will influence the success of a peptide synthesis. The peptides synthesised here have no amino acids known to be particularly troublesome (such as histidine), and no strong β-sheet regions are included, which are often the source of aggregation-related synthesis failures [171].

The development of automated microwave-assisted SPPS [166; 162] has greatly simplified peptide production, making it accessible to researchers such as ourselves with less experience in chemical synthesis.

The amino ends of the amino acids used are all protected by an Fmoc group, so that two (or more) of a given amino acid don't bond to each other, leading to an incorrect sequence on the resin. Once each coupling reaction has occurred, any remaining uncoupled amino acid molecules are washed away. The Fmoc groups of the resin-attached amino acids can then be removed and washed away. Washing is important, both to remove free Fmoc groups from the system so they cannot reattach, and to remove the deprotection chemicals so they cannot deprotect subsequent amino acids too early. The next amino acid can then be added, coupled, and deprotected, and the process is repeated until the sequence has been produced. Reactive side chains are protected during synthesis to prevent attachment to other amino acid side chains, allowing a linear peptide backbone to be formed. These protecting groups (orthogonal to Fmoc and hence not removed by the deprotection solution used during synthesis) are removed after synthesis, alongside cleavage of the peptides from the resin.

Details of the syntheses for this project are given in Appendix D.1.

2.1.2 High Performance Liquid Chromatography (HPLC)

Chromatography involves separating out different molecules according to one of their properties, such as size, hydrophobicity or charge. We use reverse phase HPLC to purify peptides according to their hydrophobicity.

Reversed phase chromatography uses a hydrophobic solid/stationary phase and a hydrophilic liquid phase [172], while normal phase uses the opposite. The peptides are dissolved, and injected onto a column while a predominantly water liquid phase flows through the system. The peptides bind to the column to reduce their exposure to the polar solvent, and non-polar solvents are needed to remove them [173]. The liquid phase is constantly flowed through the system, and a hydrophobicity gradient is carried out, to a solution of decreased polarity (in our case, predominantly acetonitrile, with a small percentage of water remaining). As the gradient is carried out, more and more hydrophobic molecules elute from the column, and can be collected. A UV detector is used to observe when the peptides are eluted, seen as a series of absorption peaks, and the separate products can be collected. Figure 2.2 shows an example of a HPLC spectrum. Similar peptides often have similar hydrophobicity, and hence separation of these can sometimes be difficult if the resolution of peaks is low. The peptides in this project were all purified via HPLC after

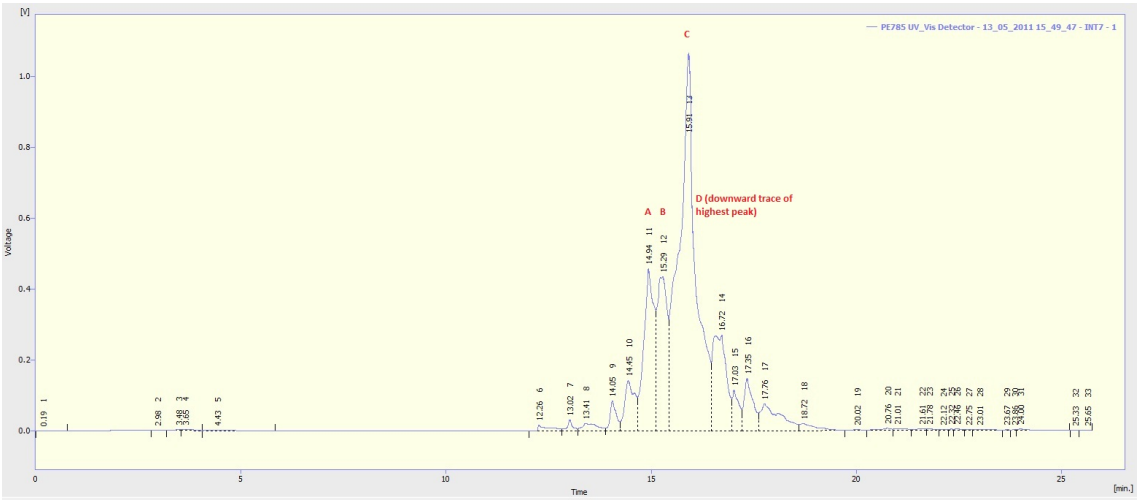


Figure 2.2: The signal received from the UV-Vis detector during a HPLC run of peptide 4 (Chapter 3). The three peaks labelled A, B, and C, and the recession of the highest peak (D) were tested using MALDI-ToF. C was identified as the target peptide, while B and D contained similar peptides, but with an additional unwanted amino acid.

synthesis. HPLC was also used at several stages in the production of disulphide-bonded

peptides, and both before and after the addition of azobenzene to peptide $\epsilon 2$.

2.2 Matrix-Assisted Laser Desorption/Ionisation Time-of-Flight Mass Spectrometry (MALDI-ToF)

MALDI was first demonstrated in 1988 by Karas and Hillenkamp, as a means of using UV laser desorption to determine molecular weight [174]. It is popular for peptides and other biomolecules of this size range, which are well above the mass region affected by matrix-related byproducts [175]. Time-of-flight (ToF) experiments involve the acceleration of ions to a specific kinetic energy (using an electric potential) and then looking at their separation after they pass through a region without field and collide with a detector. Ions of the same charge z will separate due to their varying velocities v, which are dependent (as they all have the same kinetic energy) on their masses m [176]:

$$v \propto (m/z)^{-1/2} \tag{2.1}$$

$$ToF \propto (m/z)^{+1/2} \tag{2.2}$$

MALDI relates to the fact that the ions are produced by a laser hitting a sample comprising of peptide combined with a matrix. The spectra produced give intensity against m/z, and though predominantly a singly charged peak appears [177], doubly (and more highly) charged peaks are also sometimes seen.

A successful synthesis would give a MALDI-ToF spectrum dominated by the correct product mass, with perhaps a few other additions or deletions present in small quantities. It was used in this work qualitatively to confirm the presence of the correct peptide after synthesis - if there is a molecule corresponding to the mass of the peptide synthesised in the MALDI-ToF results, it was be taken that the peptide was synthesised successfully. Of course, there are many peptides of the same mass. If one rearranged a sequence with the same residues, it would have the same mass, but if there has been a problem in the synthesis, there would be a very small chance of there being an alternative peptide of the same mass present without peptides of other masses also being present. MALDI-ToF can also indicate when a peptide has been incompletely deprotected (through the presence of masses equivalent to the peptide plus deprotecting groups). In this project we therefore tested peptides by MALDI-ToF after synthesis, to confirm both successful synthesis

and cleavage, and after HPLC to identify purified fractions of the correct peptide. The resolution of the MALDI technique varies, and can provide a far greater resolution than needed for our experiments; the systems used here for MALDI-ToF analysis are sufficient to resolve mass differences of ∼1 Da, and hence can easily resolve peptides with mass differences of ∼100 Da such as ours, or different products of peptide syntheses containing additions and deletions from the desired sequence as well as the main product. As shown in Figure 2.3, it is possible to resolve peptides and their sodiated counterparts; the spectra often contain peaks at 22 and 38 mass units higher than the expected peptide peak value due to sodium and potassium adduct ions present with the peptides. Figure 2.3 shows an example of a MALDI-ToF spectrum, where the major peak is the desired product, and the next (although small) peak present in the region is the peptide+sodium ion peak.

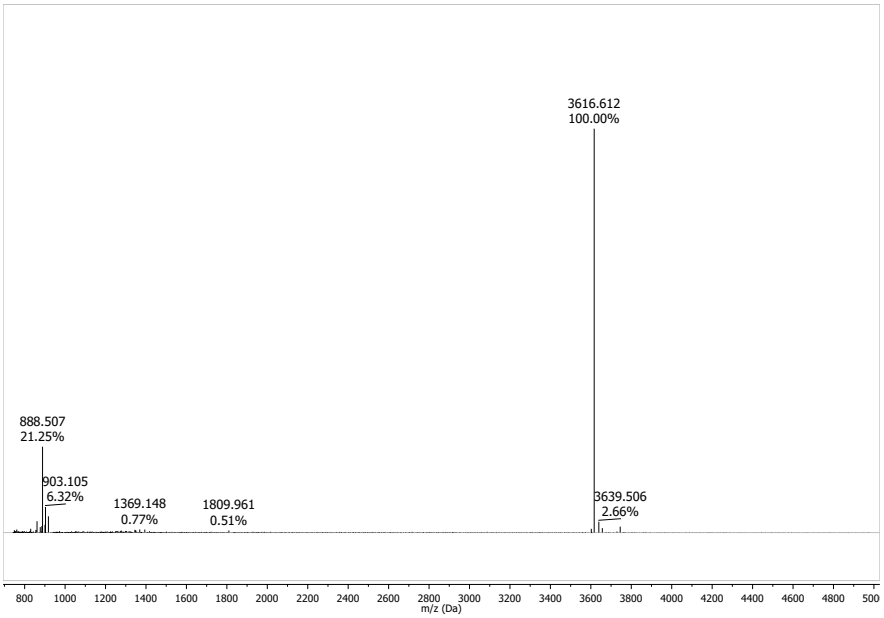


Figure 2.3: A MALDI-ToF spectrum for a fraction collected from HPLC purification, for peptide 3 of the Tumbleweed system. The spectrum shows the peptide, and a small peak due to the commonly occurring sodium adduct, 22 mass units above the peptide peak.

2.3 Characterisation Techniques

After synthesis and purification, the peptides were characterised. Circular dichroism (CD) was used to measure secondary structure, and to test thermal stability. Absorption spectroscopy, in addition to being the detection technique in HPLC, was used to calculate pep-

tide concentrations, and as an indicator of structural changes in conformational switching experiments. Dynamic Light Scattering was used to estimate sizes of peptide species. Analytical Ultra Centrifugation data carried out by EHCB has been used to complement this PhD data, as an assessment of species masses, to deduce the oligomeric states present.

2.3.1 UV-Vis Spectroscopy

Concentration

UV-Vis spectroscopy is regularly used as a measure of concentration for peptides and proteins. Each of the peptides in this project contains a Tyrosine residue. Tyrosine has a strong UV absorption around 276 nm, with an extinction coefficient of 1280 $cm^{-1}M^{-1}$ [135]. Tryptophan and Phenylalanine, the other aromatic amino acids, also have an absorption in this region which could be used if present. If peptides do not have aromatic residues, then the peptide backbone absorption at 220 nm can be used. Samples must be of a concentration high enough to give a significant aromatic absorption over background absorptions but not so high that they are subject to adverse scattering effects. Peptides with other absorbing moieties such as azobenzene can cause issues, as it is necessary that an absorption peak is attributable to one moiety; the signal should return to a baseline either side of the peak. If there are too many overlaps, or if the signal of the Tyrosine residue is swamped by another species, then another absorption peak, or a different method of determining concentration must be used.

The Beer-Lambert law can be used to calculate the concentrations of peptide solutions from absorption spectra [178]:

$$A = \epsilon c l \tag{2.3}$$

where A is absorbance, c is concentration, l is path length and ϵ is extinction coefficient.

Structural Changes

In addition to concentration calculations, studying the temporal evolution of UV absorption spectra over time can be used to track structural changes. If one measures the UV spectrum of a peptide and then induces a structural change, a change in the UV spectrum is additional evidence of a change in the molecule's environment. If there is then migration of this spectrum back towards the initial spectrum, one can use this to estimate the lifetime of the switch, or if a reverse does not occur, suggests that the switch is either long

lasting or irreversible without external intervention. The UV spectrum of ϵ2CL (Chapter 5) was monitored before and after irradiation, to confirm that a change had occurred, and to study its return to its original state.

2.3.2 Circular Dichroism (CD)

Circular dichroism (CD) is a property exhibited by chiral molecules. Molecules are chiral if their mirror images can not be superimposed over each other. Amino acids are all chiral (except for glycine), and hence all peptides except polyglycine are chiral.

Circular dichroism occurs because the electric field vector of circularly polarised light (CPL) traces out a helix in the direction of propagation, and, as photon absorption causes electron rearrangement (to higher energy states), in chiral molecules (no plane of reflection), this rearrangement also has a helical component [179]. The two circular polarisations therefore have a differing effect; chiral molecules absorb left (L-) and right (R-) circularly polarised light to different extents, and the difference in these absorptions is circular dichroism:

$$CD = \Delta A = A_{\text{L-CPL}} - A_{\text{R-CPL}} \tag{2.4}$$

This difference can be measured over a range of wavelengths. The CD spectra of peptides are taken in the UV region, which involves backbone transitions (as well as some side chain transitions). There have been found to be characteristic features of the CD spectra for particular peptide secondary structural motifs (Figure 2.4). Spectra of α-helices have two minima; one at the lowest energy transition in the peptide chromophore, the n $\rightarrow \pi^*$ transition, at 222 nm for α-helices, and one which is part of the $\pi \rightarrow \pi^*$ transition, at 208 nm (this transition has a splitting for α-helices). A fully folded helical peptide has -33,000 degcm^{2}dmol^{-1}res^{-1} at 222 nm [180], and this negative minima disappears as the peptide loses or changes its secondary structure. CD is thus frequently used to study peptide and proteins, to assess the secondary structures present, compare the structures of different peptides, and study stability and changes in structure with changing conditions (such as pH or temperature). Another method commonly used to look at peptide secondary structure is X-ray scattering, but crystallization itself can cause changes in the structures of peptides/proteins [179; 181], and hence may not represent the system as it is in solution. It is also often difficult to produce good quality crystals suitable for crystallography, so it may be more reasonable to use CD if one wishes to look at a system in solution.

It is common for CD data to be analysed by comparison with 'standards'; analysis typ-

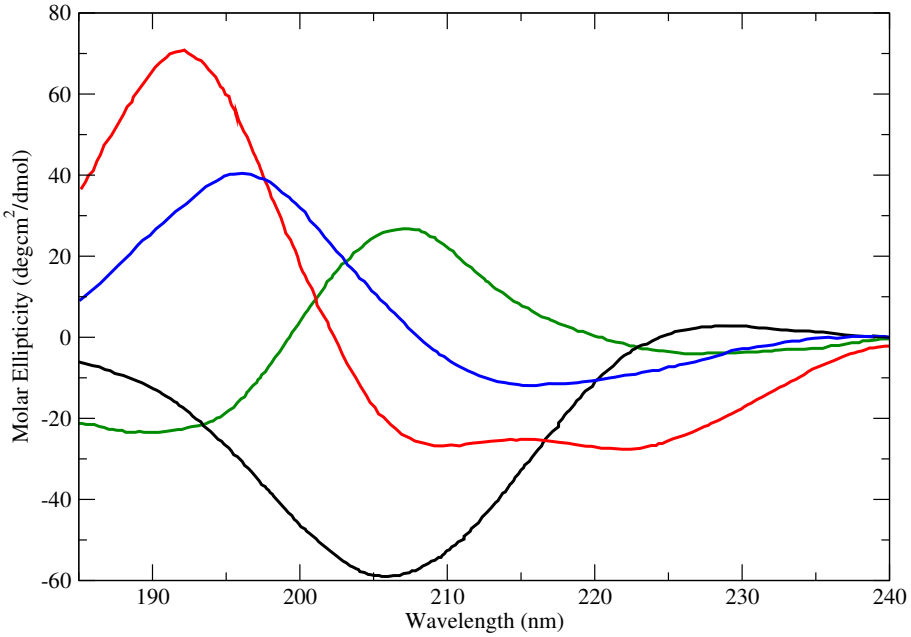


Figure 2.4: Typical CD spectra of common peptide secondary structures. α-helices (red) generally have 2 minima, at 208 and 222 nm, while β-sheets (blue) have one minima at around 216 nm [179]. Also shown are a β-turn (green) and poly-L-proline, a left handed helix found in collagen (black). Adapted from [182].

ically involves using spectra of known helical, beta sheet and random coil molecules to estimate the amount of each of these secondary structures present in a molecule, by combining the spectra of different secondary structures in different ratios to find a 'best fit' to the new molecule's spectrum. There are many packages available for estimating the percentages of each secondary structure (α-helix, β-sheet, random coil) present in samples [183; 184; 185; 186], which generally use a basis set of peptides with known secondary structure (often from X-ray studies), to compare with data. Obvious issues relating to the validity of this are whether one fully folded helical peptide will have the same spectrum as another, and whether the structures found from X-ray data are identical to the structures found in solution. Libraries are being improved and expanded over time, and fitting packages are estimating structures more and more accurately [186]. Generally, α-helices are more accurately predicted than other secondary structures [187]. In this thesis (both Chapters 3 and 5), I predominantly use direct comparison of my own samples as a means of interpreting how folded peptides are, and whether helicity is increased in mixtures as a result of coiled-coil interactions. In both types of comparison, it is important that gases and other molecules (such as the solvent system) involved in the system do not cause

absorptions of their own (or that these are minimised) so that the results seen are representative of the peptides, and not other presences. The spectropolarimeter is purged of oxygen, as it absorbs strongly in the 200 - 10 nm range [188] (and the UV produced by the lamp can also react with surrounding oxygen to form ozone, highly undesirable), prior to use, and solvents which do not absorb in the regions of interest are chosen. It is also important to keep conditions constant between samples (temperature, pH, solvent). By looking at individual peptides, and peptides mixed with others, one can confirm an interaction between them by whether the mixture's measured spectrum differs from what is predicted by combining two peptides' individual spectra.

2.3.3 Dynamic Light Scattering (DLS)

Dynamic Light Scattering uses the scattering of light by molecules in solution to probe their properties. The experimental setup is shown in Figure 2.5. The intensity of scattered light will vary over time as it is scattered by particles which are moving with Brownian motion. Fluctuations in the intensity are measured, and used to calculate the autocorrelation function - a measure of correlation between intensity at a time t, and a short time later t+τ:

$$G(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} I(t)I(t+\tau)dt \tag{2.5}$$

The correlation between two signals at times very close to each other will be high, as the scattering particles will not have had time to move significantly, but those signals separated by long times will have no correlation in a solution undergoing a random process like Brownian motion. Therefore, for a single species of particle, the correlation function decays with an exponential form [189]:

$$G(\tau) = A + Be^{-2Dq^2\tau} \tag{2.6}$$

where A accounts for a baseline, B accounts for the fact that the resolution of fluctuations seen by the detector depends on the angle the detector subtends (if above a certain angle) relative to the separation of scattering molecules and D is the diffusion coefficient. q is the scattering vector [190]:

$$q = \frac{4\pi n}{\lambda} \sin\left(\frac{\theta}{2}\right) \tag{2.7}$$

where θ is the scattering angle and n is the refractive index of the solution.
Smaller particles will move more quickly than larger ones, and hence their correlation will decay more quickly.

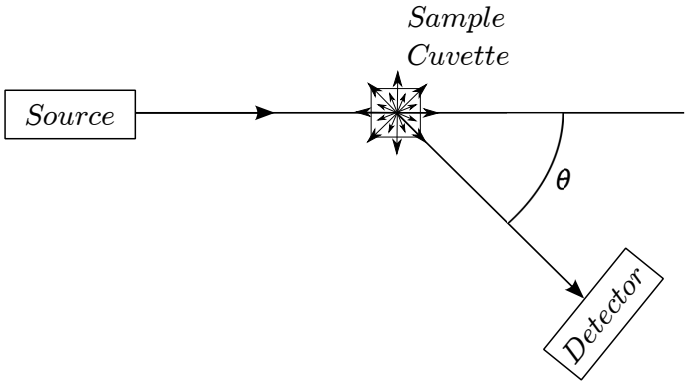


Figure 2.5: DLS experiment: light, usually from a laser, is passed through a sample solution, and light scattered in one direction is collected. The fluctuations in the intensity of the light are related to the properties of the solute, and information such as the diffusion coefficient, and hence the hydrodynamic radius, can be extracted.

Fitting the measured autocorrelation function to a theoretical function of this form, and ensuring the suitability of a single exponential form (used for monodisperse molecules) by calculation of χ^2, the diffusion coefficient D can be determined, and then the hydrodynamic radius r found using the Stokes-Einstein equation [191]:

$$D = \frac{k_B T}{6\pi\eta r} \tag{2.8}$$

where T is the temperature, k_B is the Boltzmann constant and η is the viscosity. The hydrodynamic diameter is the diameter of a sphere diffusing at the same rate as the measured species. We expect coiled-coil peptides to be around 1 nm long per heptad, and so would expect 3-4 nm structures for coiled coils formed by the hub peptides. As coiled coils are more cylindrical than spherical in nature, we would expect values of hydrodynamic diameter to vary slightly from the true length of the coiled coils, but they will give an indication of whether a structure in the 3-4nm region (for a single coiled coil such as 1,2) is forming. If we see values much smaller than this, then the peptides are likely forming compact structures, and if larger, fibrous structures. For longer peptides, formation of a hairpin and subsequent opening of the hairpin should result in a significant size change seen between DLS measurements. DLS measurements have been taken for both the systems in Chapters 3 and 5.

2.3.4 Analytical Ultracentrifugation (AUC)

AUC involves the application of centrifugal force to solutions, and observation of the redistribution of the solute. An experimental setup is shown in Figure 2.6. The main types of AUC experiment are Sedimentation Velocity (SV) and Sedimentation Equilibrium (SE). Both require measurements at several concentrations, and for SE, several rotation speeds. Only SE results are discussed in this thesis.

Theory

When samples are centrifuged, they are subjected to centrifugal force [192]:

$$F = m\omega^2 r \tag{2.9}$$

The effective mass of the particles (here, peptides) is slightly lower than their non-solvated mass, due to upthrust in response to their displacement of solvent:

$$m_{\text{effective}} = m - m\bar{\nu}\rho_{\text{solvent}} \tag{2.10}$$

where $\bar{\nu}$ is the partial specific volume.

In SE, rotor speeds are chosen to allow competition between sedimentation and opposing diffusion, acting to reduce the concentration gradient being produced. An equilibrium is established, where a time-independent concentration distribution is seen, and from this, molar mass can be measured [193; 194]. A typical concentration distribution for a sedimentation equilibrium experiment is shown in Figure 2.7. The concentration profile of a single species is given by Equation 2.11.

$$C(r) = C_0 e^{m_{\text{eff}}\omega^2(r^2 - R_0^2)/2k_B T} \tag{2.11}$$

Analytical Ultracentrifugation is used in studies such as these to elucidate whether dimers or higher order oligomers are being formed, through measurement of the masses of the present species.

2.4 Summary

Combining all of these techniques, we can prove that we have interacting peptides, which are forming structures of a given mass and size. Further to this, experiments to prove the exclusivity of the peptides are discussed in Chapter 3, and simulations are described and analysed in Chapter 4.

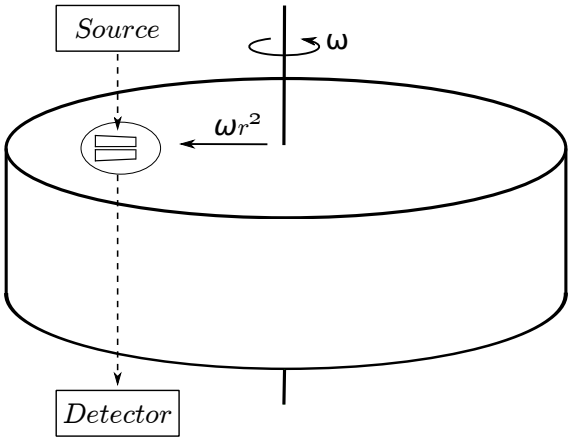


Figure 2.6: AUC experimental setup: a sample and standard are centrifuged, and observed using a detection system.

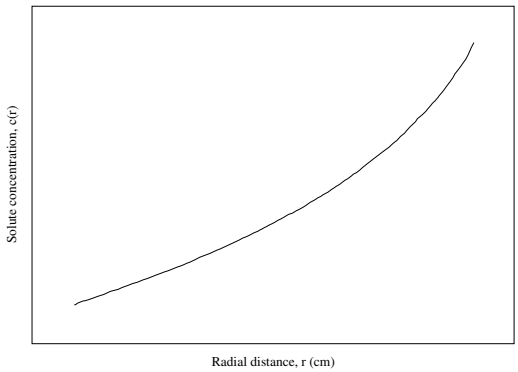


Figure 2.7: A typical spectrum seen for the concentration of a sample, as a function of radius, for SE. Absorbance (via the Beer-Lambert law) directly correlates with concentration. c(r) at equilibrium for a single species is shown.

Chapter 3

The Tumbleweed Hub

3.1 Hub Structure

As discussed in Section 1.5.4, peptides 1, 2, 3, 4, 5 and 6 are the components of the Y hub. The initial aim of this PhD project was to synthesise and characterise these peptides. This chapter initially describes the characterisation of individual and mixtures of single hub peptides, without any covalent bonding between them. As discussed in Section 1.5.4, to form the full hub, peptides from different coiled coils have to be chemically bonded, (from N to C termini) 6 to 1, 3 to 2 and 4 to 5, and when these three chemically bonded pairs are mixed, the formation of the hub will be completed by preferential coiled-coil formation between 1 and 2, 3 and 4, and 5 and 6 (Figure 3.1). In this thesis, the covalent bonding between 1 and 6, 2 and 3, and 4 and 5 has been via disulphide bonds between cysteine residues, positioned at one end of each hub peptide (Table 1.1). Throughout this chapter, peptide pairs denoted 'a,b' are individual peptides *a* and *b* mixed together, while 'a-b' denotes *a* and *b* linked via a disulphide bond. Through the characterisation of the individual and mixtures of peptides, the interactions between them are investigated. The specificity of the designed coiled coils, which is vital for a stable hub with three different feet to form, is also covered in this chapter.

3.2 Experimental Aims

For this system:

1. All of the six hub peptides were synthesised (four by LSRS, and peptides 1 and 6 by Marc Bruning).

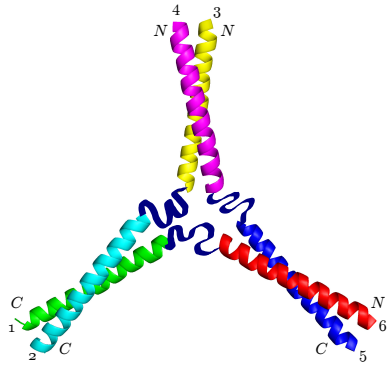


Figure 3.1: The 'Y' configuration of the Tumbleweed hub, with termini labelled, showing the antiparallel nature of 5,6 and the parallel coiled coils 3,4 and 1,2. To form a hub, coiled-coil pairs (1,2), (3,4) and (5,6) need to preferentially interact, while 1 and 6, 2 and 3, and 4 and 5 are forcibly linked to ensure all of the coiled coils are connected to form one structure.

2. All peptides were purified via HPLC.
3. The CD spectra of each peptide and pair of peptides was measured.
4. DLS measurements were taken for each of the designed coiled-coil pairs.
5. Disulphide bonds were formed between pairs of peptides.
6. Disulphide bonds in the system were allowed to rearrange to investigate pair specificity of the coiled-coil peptides.

In addition to these experiments, data taken by Elizabeth Bromley (EHCB) on this system is used in the discussion, and labelled appropriately.

3.3 Results and Discussion: Individual Peptides

3.3.1 CD

In this work I have used CD to study the helicity of the hub peptides, and the changes they undergo when mixed with their designed coiled-coil partners and other peptides in the Tumbleweed system.

In addition to a CD spectrum taken at 20 $^{\circ}$ C (shown in Figure 3.2), studies of the stability of the peptides were also carried out, by measuring their CD at 222 nm as a function of temperature, between 10 and 90 $^{\circ}$ C (Figure 3.3). For individual peptides,

circular dichroism measurements were performed on 20 μM solutions in PBS (pH 7.4), to assess their helicity without a partnering peptide. CD spectra were also measured for

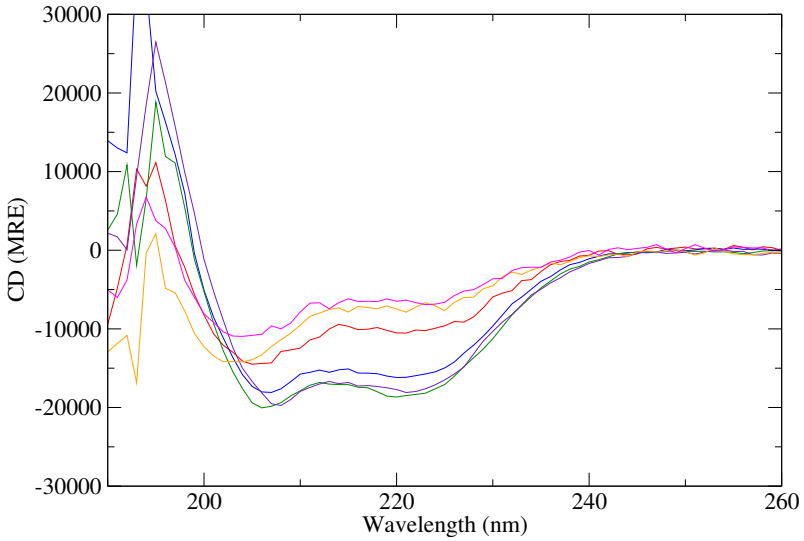


Figure 3.2: Individual CD spectra for the six Y hub peptides (1 to 6), each measured at 20 μM concentration in PBS buffer (pH 7.4), at 20°C. The peptides 1 to 6 are shown in red, blue, green, purple, orange and magenta respectively.

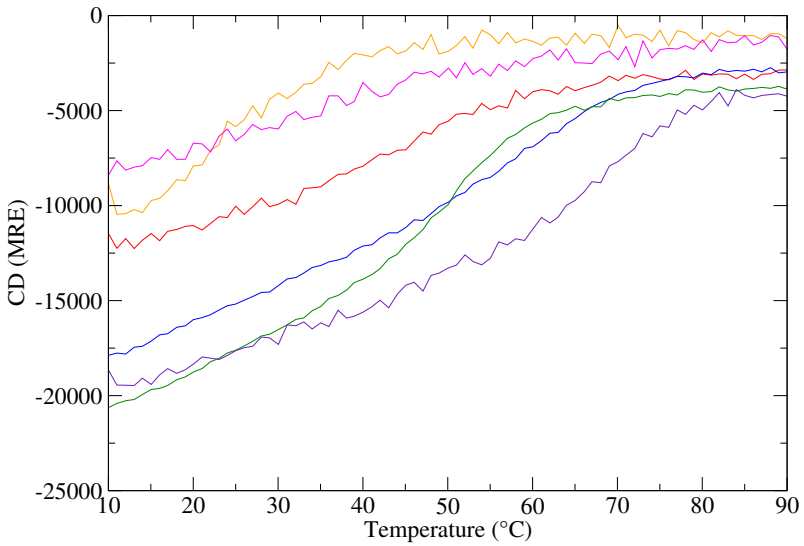
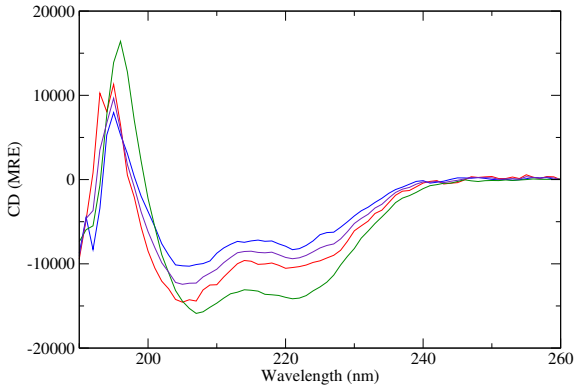


Figure 3.3: Melt spectra for the six Y hub peptides (1 to 6), each measured at 20 μM concentration in PBS buffer (pH 7.4), from 10 to 90°C. The peptides 1 to 6 are shown in red, blue, green, purple, orange and magenta respectively.

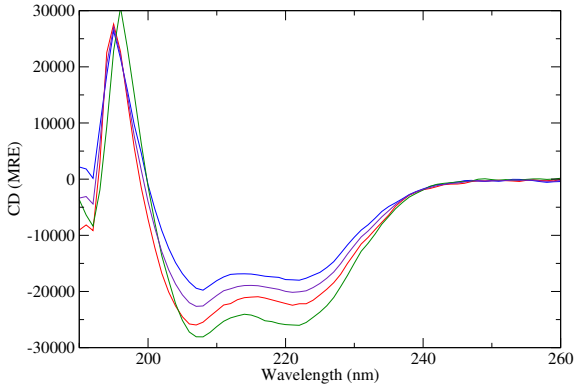
each of the 15 possible two-peptide mixtures. Each peptide in the mixtures was at a concentration of 20 μM, and hence the mixtures had a total peptide concentration of 40

μM. All of the spectra have been converted to Molar Residue Ellipticity, so as to remove the effect of varying residue concentration, as the peptides differ in sequence length.
For each two-peptide mixture, the individual peptide spectra have been plotted alongside the mixture data, and a theoretical mixture spectrum calculated from the single peptide data. These theoretical spectra are what would be expected if there was no interaction between the two peptides present, a weighted average of the two individual peptides' CD spectra (Appendix F.2.3). Figure 3.4 shows the CD data for the designed dimers, while the other 12 mixtures are shown in Figure 3.5. Melt data for the 15 mixtures was also taken; the designed pair data is shown in Figure 3.6 (the other 12 spectra are in Appendix E.2). The resultant melting temperature values are shown in Table 3.1. Finally, the CD spectrum of all six peptides mixed together was taken, and compared with the spectra predicted from the six individual peptides (no hetero-oligomeric interactions), and from the three designed pairs (Figure 3.7). The melt data of the full set of peptides was also taken and compared with the sum of the designed pairs (Figure 3.8).

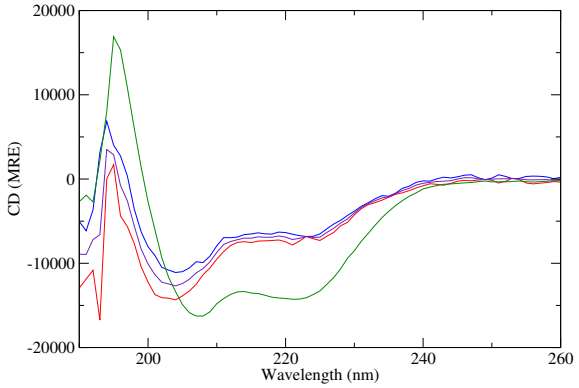
The design requisite for the peptides in the Tumbleweed hub was not to form individual α-helices, but to form helical structures - dimeric coiled coils - in the presence of a specific partner. There may be some individual helicity, due to an overlap between the design of a coiled-coil peptide pair and a helical peptide, but there should be greater helical structure when the peptides are in a 1:1 mixture with their intended partners than when they are alone in solution. To form the hub, they must also preferentially bind to this partner when given a choice between the partner and other hub peptides.
As shown in Figure 3.2, the individual peptides, as indicated by their CD spectra, have varying degrees of helicity. Their helicity could be due to the peptide being intrinsically helical, or because multiple copies of the peptide are forming homo-oligomers.
The CD designed pair data (Figure 3.4) shows that all three fulfil one element of the design criteria; they are well designed to form coiled coils. There is a consistent increase in helicity when the pairs are mixed together. The measured mixtures show significantly increased helicity than the theoretical spectra, which would be expected if the peptides present were not interacting. In all three of the pairings, the mixture is also more helical than the individual peptides.
An unsurprising but complicating factor in testing the 15 peptide mixtures is evident from the other 12 possible pairings (Figure 3.5). All of the non-designed mixtures also appear to interact and form hetero-oligomeric species that are more helical than their homo-



(a) 1,2



(b) 3,4

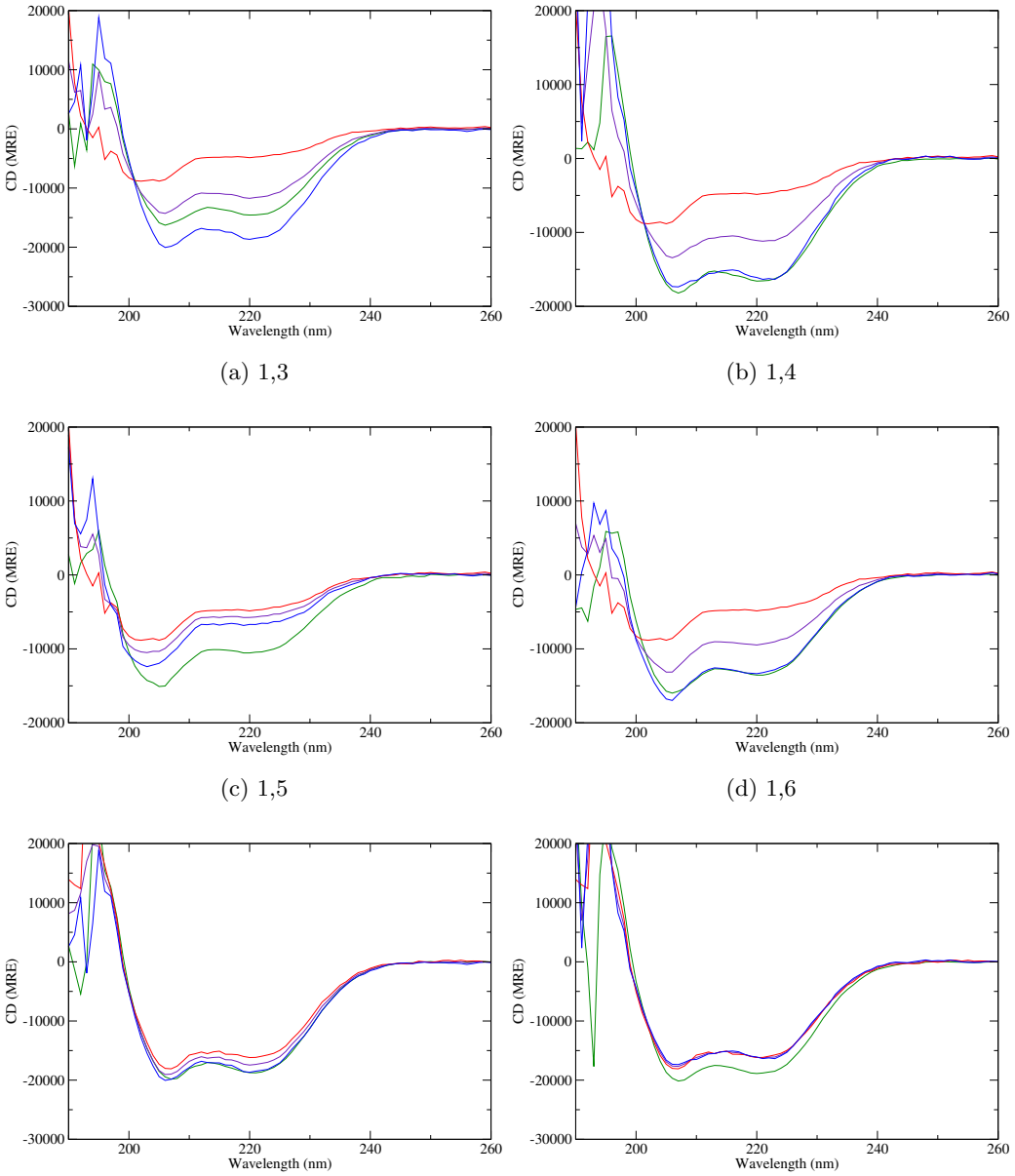


(c) 5,6

Figure 3.4: CD spectra of the three mixtures of the designed pairs of peptides - 1 and 2, 3 and 4, 5 and 6 (shown in green). Each figure also shows the individual peptide spectra (red and blue, in numerical order), and the theoretical signal for a mixture of the two without interaction (a combination of the two single peptide signals, shown in purple). Each sample had a concentration of 20 μM per peptide, in PBS buffer (pH 7.4), measured at 20°C.

Table 3.1: Melt temperatures of peptides studied

Peptide Sample	Type	Melting Temperature (°C)
1	single	49 ± 3
2	single	65.6 ± 0.2
3	single	52.8 ± 0.1
4	single	75 ± 1
5	single	-
6	single	43 ± 3
1,2	designed pair	52.2 ± 0.4
3,4	designed pair	>80
5,6	designed pair	>80
1,3	not designed pair	50.1 ± 0.1
1,4	not designed pair	69.2 ± 0.2
1,5	not designed pair	64 ± 2
1,6	not designed pair	50 ± 3
2,3	not designed pair	69.8 ± 0.3
2,4	not designed pair	73.4 ± 0.4
2,5	not designed pair	70.5 ± 0.5
2,6	not designed pair	65.3 ± 0.2
3,5	not designed pair	57.8 ±0.4
3,6	not designed pair	62.2 ± 0.1
4,5	not designed pair	67.4 ± 0.3
4,6	not designed pair	73.2 ± 0.3
1,2,3,4,5,6	full mixture	>80



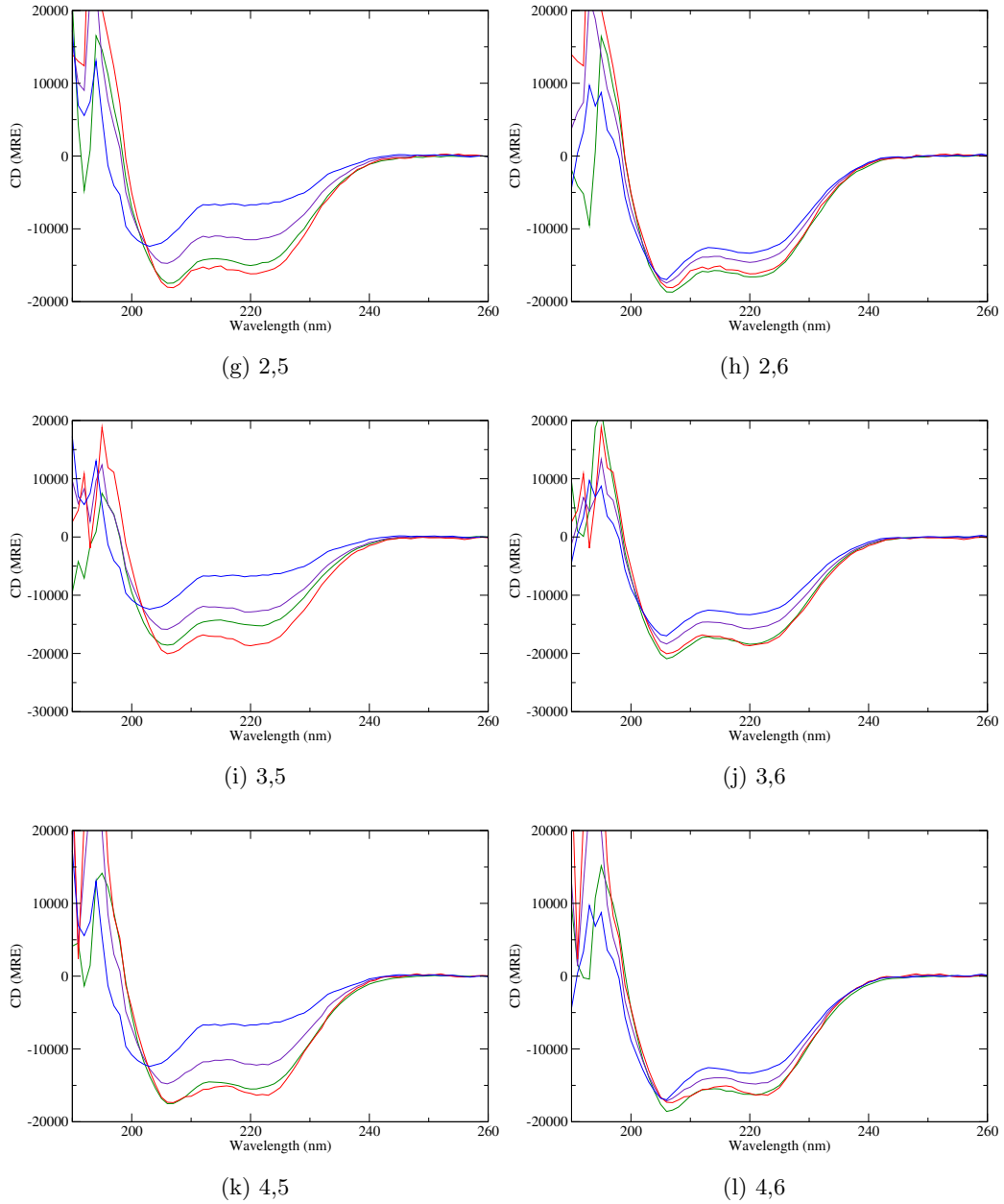
(a) 1,3 (b) 1,4

(c) 1,5 (d) 1,6

(e) 2,3 (f) 2,4

Figure 3.5: CD spectra of the 12 non-designed peptide pairings. Each sample had a concentration of 20 μM per peptide, in PBS buffer (pH 7.4), measured at 20°C. The spectra of the individual peptides present in the mixture are shown in red and blue respectively, while the spectrum of the mixture is shown in green. The theoretical spectrum for the mixture, calculated from the individual spectra, is shown in purple.

oligomeric species. As all of the peptides are designed to have a hydrophobic seam, it is no surprise that interacting with other peptides, or copies of themselves, to bury hydrophobic residues, is energetically favourable over remaining as monomers in solution. All 15 of the mixtures appear to have melting temperatures sufficiently high to be functional at room



(g) 2,5 (h) 2,6

(i) 3,5 (j) 3,6

(k) 4,5 (l) 4,6

Figure 3.5: CD spectra of the 12 non-designed peptide pairings. Each sample had a concentration of 20 μM per peptide, in PBS buffer (pH 7.4), measured at 20°C. The spectra of the individual peptides present in the mixture are shown in red and blue respectively, while the spectrum of the mixture is shown in green. The theoretical spectrum for the mixture, calculated from the individual spectra, is shown in purple.

temperature, in which future experiments could be expected to be performed in. With this interaction between non-designed pairs of peptides visible, it is even more important to use other methods to confirm that the set of six follow the design criteria of forming the three designed dimers when they are all mixed, and not a mixture of all 15 possibilities

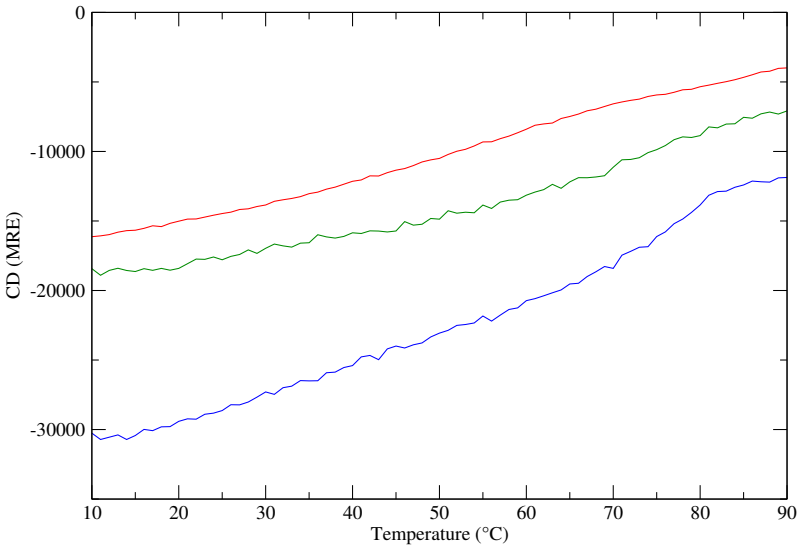


Figure 3.6: Melt curves for the three designed dimers. Each sample had a concentration of 20 μM per peptide, in PBS buffer (pH 7.4), measured from 10 to 90°C. 1,2 is shown in red, 3,4 in blue and 5,6 in green.

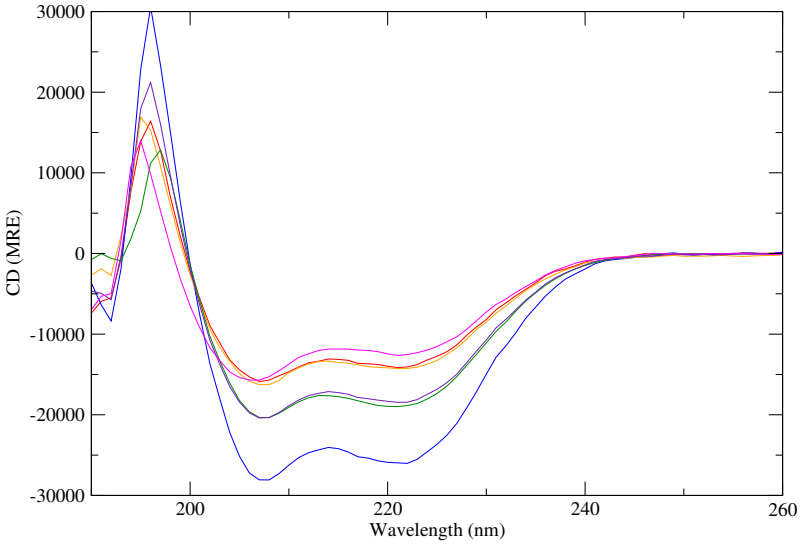


Figure 3.7: CD spectrum of all six hub peptides mixed together (green), compared with the spectrum expected from the individual spectra (i.e. if there was no inter-peptide interaction, shown in magenta), and from the designed pair spectra (i.e. if the designed pairs, and those only, interact, shown in purple). The spectra for 1,2 is also shown, in red, 3,4 in blue and 5,6 in orange. The sample had a concentration of 20 μM per peptide, in PBS buffer (pH 7.4), measured at 20°C.

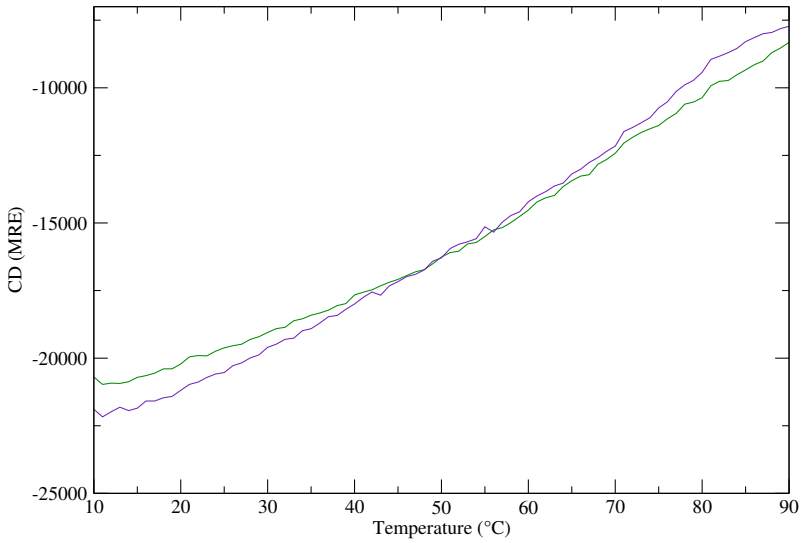


Figure 3.8: Melt spectrum of all six hub peptides mixed together (shown in green), compared with the spectrum expected (shown in purple) from the designed pair melt spectra (i.e. if the designed pairs, and those only, interact). The sample had a concentration of 20 μM per peptide, in PBS buffer (pH 7.4), measured from 10 to 90°C.

(or an even greater number of higher order oligomeric species).

The data from the mixture of all six peptides is positive towards this conclusion; the measured CD spectrum and melt are very similar to what would be expected for a mixture of the three designed pairings, as calculated using the spectra from the three individual designed pairs (Figures 3.7 and 3.8).

3.3.2 DLS

DLS was used to measure the hydrodynamic diameters of the three designed coiled-coil peptide pairs, at concentrations of 100 μM per peptide. 50 measurements were taken of each pair (with each measurement being the result of five 10-second DLS runs), and peaks in the region of interest, of intensity as a function of size, on the DLS spectra were fitted with a Gaussian function, to extract the peak values. An example of the fits is shown in Figure 3.9. Using a Gaussian fit should be more accurate than using the point of maximum intensity, as this could be biased to one side of the peak due to a low number of data points defining the peaks' maxima. An average was then found for each pair of peptides. Some of the spectra had more than one peak in the region of interest (Figure 3.9), but as these were likely artifacts due to the algorithm used by the DLS software to create the spectra, as opposed to being true second species, they were discarded and not

included in the averaging process. Most spectra also had peaks in the 10^2-10^4 nm region of the spectrum, and a few had peaks in the 40-100 nm region, but not a constant alternative species present. To decide a suitable numerical cutoff to remove bad fits, the standard deviation of each data set was calculated (using normalised intensities, as some runs had higher intensities than others), for the 0-20 nm region. It was found that a cut off of 0.08 (any sets with a standard deviation above were discarded) removed any visibly bad fits to peaks, but that this did not necessarily cut out all of the double-peaked sets, as some of the second peaks were small, and the first peaks were fit well to a Gaussian function, making the overall fit better than some fits to single peaked spectra. The average size from each data set, with no exclusions, all double peaked data sets removed, all above 0.08 standard deviation data sets removed, and both above-0.08 standard deviation sets and double peaked sets removed, are shown in Table 3.2. It is obvious from the small differences between the results in Table 3.2 that the exclusions do not greatly affect the resulting averages.

1,2

All 50 data sets had peaks in the region of interest, but only 48 were fitted, as the remaining two had two peaks of competing size, making the choice between them difficult for the fitting protocol, when they would be discarded later regardless, due to the twin peaks. A further 11 were discarded later due to their possessing a double peak, leaving 37 sets with a single peak. Four more data sets were excluded based on their standard deviations. 33 sets were used for the final average, of 3.41 $\pm$ 0.08 nm.

3,4

All 50 data sets contained peaks in the region of interest. Five were excluded due to the presence of a second smaller peak in the region of interest, and two more due to their standard deviation values. The average from 43 data sets was 5.04 $\pm$ 0.06 nm, consistent with the average found if the main peak from all 50 data sets was used.

5,6

Only 43 data sets had a peak in the region of interest, immediately reducing the set size. Two more were excluded by the selection criteria, giving an average from 41 sets of 4.25 $\pm$ 0.05 nm.

Table 3.2: DLS Averages: Tumbleweed Designed Pairs

Sample	**Criteria**	**Sets used**	**Average (nm)**	**Standard Deviation (nm)**	**Standard Error (nm)**
1,2	All data sets	48[1]	3.48	0.64	0.09
	Double peak sets removed	37	3.43	0.50	0.08
	$\sigma < 0.08$	34	3.43	0.49	0.08
	No double peaks, $\sigma < 0.08$	33	3.41	0.49	0.08
3,4	All data sets	50	5.01	0.42	0.06
	Double peak sets removed	45	5.05	0.42	0.06
	$\sigma < 0.08$	47	5.02	0.41	0.06
	No double peaks, $\sigma < 0.08$	43	5.04	0.42	0.06
5,6	All data sets	43[2]	4.23	0.34	0.05
	Double peak sets removed	42	4.24	0.32	0.05
	$\sigma < 0.08$	41	4.25	0.32	0.05
	No double peaks, $\sigma < 0.08$	41	4.25	0.32	0.05

[1] Two sets were removed prior to fitting, as they contained two competing peaks in the region of interest.

[2] 7 data sets had no peaks in the region of interest.

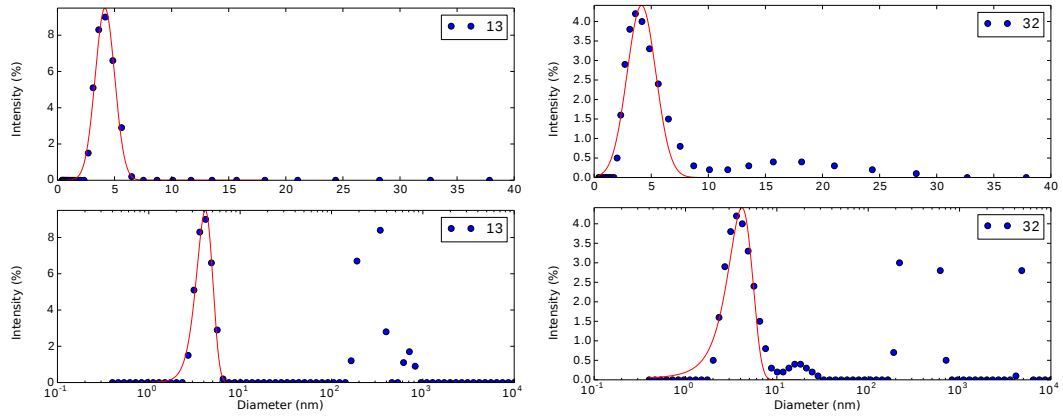


Figure 3.9: Two examples of DLS data sets (runs 13 and 32) for the 5,6 sample. The first shows a single peak, well fitted to a Gaussian function, while the second is an example of a double peak, and was not included in the averaging process.

Discussion

The DLS data adds some evidence backing the success of the design requisite of blunt ended coiled-coil dimers. As discussed in Section 2.3.3, one would expect each of these coiled coils to form a species of 3-4 nm, and hence have a hydrodynamic diameter of similar size. It is unusual that the 5,6 pair appears to have a larger hydrodynamic radius than 1,2, as both 1,2 and 3,4 are four-heptad coiled coils, while 5,6 has only three heptads. However, the data from all three pairs indicates strongly the presence of a species with a size close to that expected of single peptide-lengthed species, such as those shown in Figure 1.10. One cannot tell from DLS data the oligomeric state of the coiled coil, as a trimer or tetramer would give a similar size, but one can exclude the possibility of fibrous species. The data implies that structures of the length expected from blunt-ended coiled coils are forming. Sticky-ended peptides (Figure 1.14) would result in fibrous structures of longer lengths, and there was no persistent peak between these ∼4 nm peaks analysed and those caused by dust/other contaminants to indicate the presence of such a species.

3.3.3 AUC

The AUC data shown here was carried out on other syntheses of these peptides, by EHCB, and is included here to support this work. All samples were made up so that their concentrations corresponded to optical depths between 0.3 and 0.45. The molecular weights measured are shown in Table 3.3. The original data is in Appendix E.3. The molecular

Table 3.3: AUC Data: Designed Pairs

Sample	**Expected Molecular Weight**	**Measured Molecular Weight**	$\bar{v}$
1,2	7041	7000 ± 1000	0.7467
3,4	7397	7100 ± 1000	0.7467
5,6	5816	5900 ± 1000	0.764

weights measured by AUC for the designed peptide pairs were all consistent with their expected molecular weights, within the experimental error given. Two of the three designed pairs had measured masses within 100 Da of their expected masses, even though they each had an error of ± 1000 Da. The third, (3,4), had a measured value within 300 Da of the expected value, with an error of ± 1000 Da. This data indicates that dimers are being formed as opposed to higher order (and heavier) species.

3.4 Results and Discussion: Disulphide-bonded Peptides

3.4.1 Disulphide Bonding Between Peptides

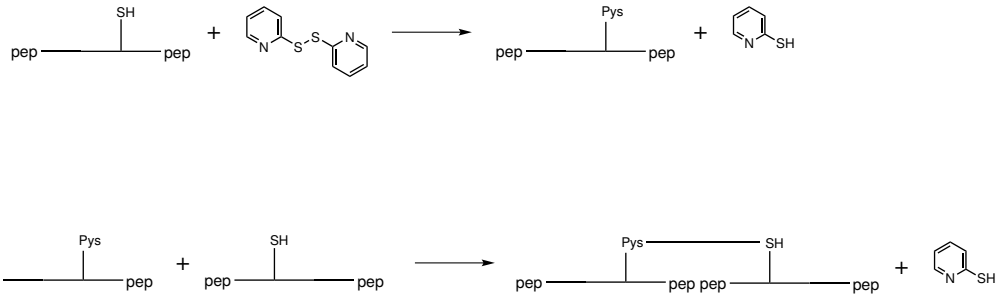


Figure 3.10: Formation of a disulphide bond between peptides, using Aldrithiol-2. One cysteine-containing peptide is first reacted with Aldrithiol-2, and then with a second cysteine-containing peptide.

One method of bonding the peptides together is via disulphide bonding, mediated by Aldrithiol-2 (also known as 2,2'-Dipyridyldisulphide and 2,2'-Dithiodipyridine), an oxidising agent. In order to facilitate this, all of the peptide sequences include a single cysteine residue. Peptides 1, 2 and 4 were each individually treated with Aldrithiol-2 to form S-Pys cysteine sidechains [195]. Each of these was then mixed with another hub peptide (6, 3, and 5 respectively) to form a disulphide bond between the cysteine residues (Figure 3.10).

The resultant bonded peptide masses are shown in Appendix C.1. The peptides were purified after each step to remove any unreacted Aldrithiol-2, and the waste molecules formed by each reaction.

3.4.2 Proving Exclusivity of Coiled-coil Pairs

In order to prove the exclusivity of the coiled-coil pairs, the use of the disulphide bonding property of cysteine can again be used. Glutathione is used as a reducing and oxidising agent for peptides/proteins, depending on which of its forms is used (Figure 3.11). In order to gain evidence that the coiled-coil hub will successfully form, we performed experiments where all three disulphide bonded species were mixed in a 1:1:1 ratio, and placed in a mildly reducing environment, sufficient that disulphide bonds could be broken and new ones reformed, but were not reduced fully, returning the original individual peptides. For this we used a low concentration of reduced glutathione. If the coiled-coil peptide pairs preferentially bind, they should become disulphide bonded when the three originally disulphide bonded pairs are given the opportunity to rearrange (Figure 3.12). The MALDI-ToF spectrum of (6-1, 2-3, 4-5) in a 1:1:1 ratio shown in Figure 3.13a was performed on a sample of the three species mixed in water pretreated with TFA, and measured within an hour of mixing. This lowering of the environment's pH (a total of 7.8μl peptide in water was placed in 50μl of pH 2 TFA-water) appears to have kept the original species intact; a test experiment, with the three species mixed in water, indicated that this environment was sufficient to allow the disulphide bonds to reform within days, and the disulphide-bonded designed pairs were seen on a comparable level to the starting peptides. The ability of the disulphide bonds to exchange so easily in water indicates that for future experiments, the conditions of the system need to be more tightly controlled to prevent rearrangement and subsequent dissociation of the hub components, and that there may be more reliable methods of ensuring the hub's stability than via disulphide bonding. The two main peaks visible are the 2-3 and 4-5 species; 6-1 is one of the smaller peaks seen, of a comparable height to some of the minor impurities seen in the starting materials. The MALDI-ToF spectrum shown in Figure 3.13b shows the state of the system when the (6-1), (2-3), and (4-5) species were mixed in PBS buffer, with a low concentration of reduced glutathione (after 5 minutes, water pretreated with TFA was added to quench the reaction). The spectrum has altered significantly. The disulphide bonds have rearranged. The peaks for the designed pairs in disulphide bonded form, 1-2, 3-4 and 5-6 are all present, and the

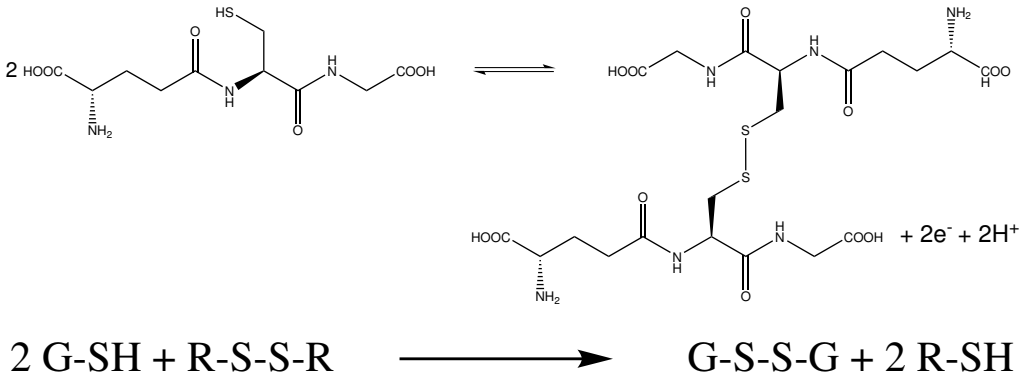


Figure 3.11: The reduced (left) and oxidised (right) states of glutathione, known as GSH (glutathione) and GSSG (glutathione disulphide) respectively. In nature, GSH undergoes thiol-disulphide exchange with protein disulphide bonds to produce GSSG. We will use the same reaction to break disulphide bonds between our peptides, but at a low enough concentration that disulphide bonds can be reformed.

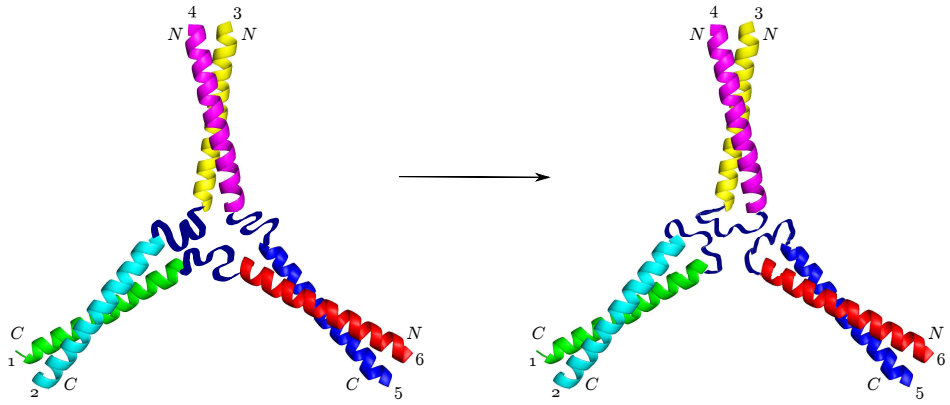
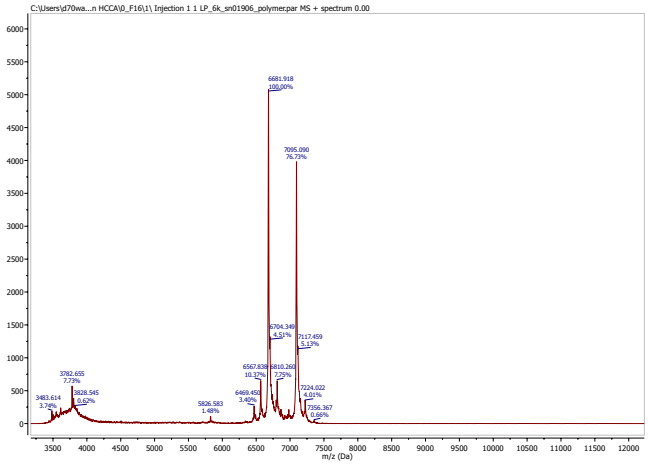
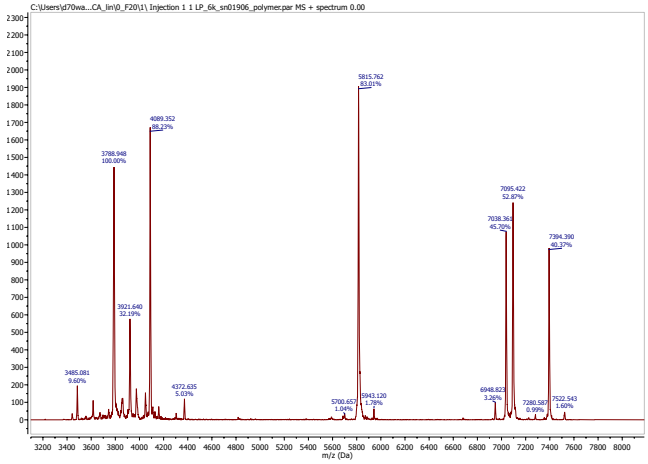


Figure 3.12: If the coiled-coil hub peptides preferentially interact, when allowed to rearrange, the disulphide bonds should switch from linking the chemically reacted pairs to between the designed pairs.

fourth strong peak belongs to the starting pair 2-3. There appears to be a trace amount of 4-5 remaining, no 6-1, and none of the other undesired pairings are present. The other small peaks are impurities. This provides rather compelling evidence that the designed pairings are highly preferred interactions in the system.



(a) Glutathione MALDI-ToF 1



(b) Glutathione MALDI-ToF 2

Figure 3.13: MALDI-ToF spectra of the species present before and after treatment of the hub with glutathione. a) The two main peaks in the spectrum correspond to the 3-2 (MW 7095) and 4-5 (MW 6682) disulphide-bonded peptides. Additionally there are peaks at the $(M + Na)^+$ masses (7117 and 6704 for 3-2 and 4-5), and the smaller peaks in this area are from the 6-1 peak (6470), and peptides which have been formed from these correct pairings of peptides, but where one of the contributing peptides had an additional/missing residue (as seen in the single peptide MALDI-ToF spectra in Appendix E). b) The dominant species present (in the region of interest) after rearrangement are 6-5 (MW 5814), 1-2 (MW 7039), 3-4 (MW 7395) and 3-2 (MW 7095).

3.4.3 Additional Data for Disulphide-bonded Peptides

The data in this section was taken by EHCB on a previous synthesis of the hub peptides. The findings of the glutathione experiments indicate that the current peptides, although

prepared in the same preparation conditions, are in an environment which does not allow long-term disulphide bonding between non-designed dimers, and hence we would not expect to see a hub-sized species in AUC or DLS data due to rearrangement of disulphide bonds and consequent dissociation of the hub components. However, in a previous synthesis, this issue was not seen, and successful data was taken, supporting the design success of the hub.

CD

The three disulphide-bonded peptides, the pairs of disulphide-bonded peptides, and the full hub (Figure 3.14) all give strong helical CD signals. The melt data for the disulphide-bonded peptides is shown in Figure E.8.

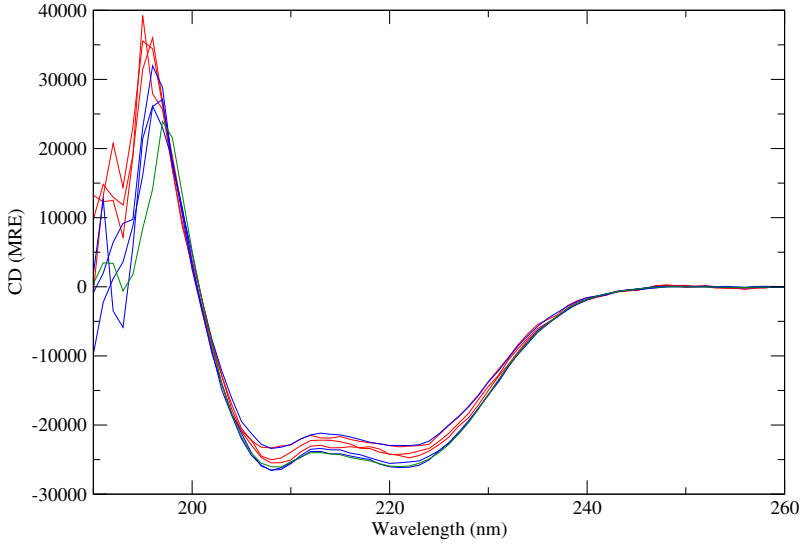


Figure 3.14: CD data of the disulphide-bonded peptides, taken by EHCB. The red spectra are the single disulphide-bonded peptides, 1-6, 2-3 and 4-5, while the blue spectra are the pairs of disulphide-bonded peptides, (1-6, 2-3), (2-3, 4-5) and (4-5, 1-6). The green spectrum is the mixture of all three disulphide-bonded peptides.

DLS

Figure 3.15 shows a representative sample of runs for each of the three disulphide-bonded peptides, 1-6, 2-3, and 4-5, the three possible pairs of disulphide-bonded peptides, and all three disulphide-bonded peptides mixed together. The three single disulphide-bonded peptides have peaks around 4 nm (two between 3 and 4 nm, and one slightly above 4 nm), while the pairs all have a higher value around 5 nm. The three disulphide-bonded

peptide mixture also has a value centred around 5 nm. No unusual behaviour is seen here. The single peptides have hydrodynamic diameters which are compatible with hairpin formation, which is not unexpected given the behaviour seen in the CD spectra of the peptide mixtures without disulphide bonds, indicating interaction between undesigned pairs in the absence of designed partners. The pairs have larger hydrodynamic diameters that the single disulphide-bonded peptides, indicating different interactions from the single species: interactions between the two species present. The hub would be expected to have a similar diameter to the pairs, as we would not expect a change in the longest dimension of the system when moving from a disulphide-bonded pair of peptides to the full hub.

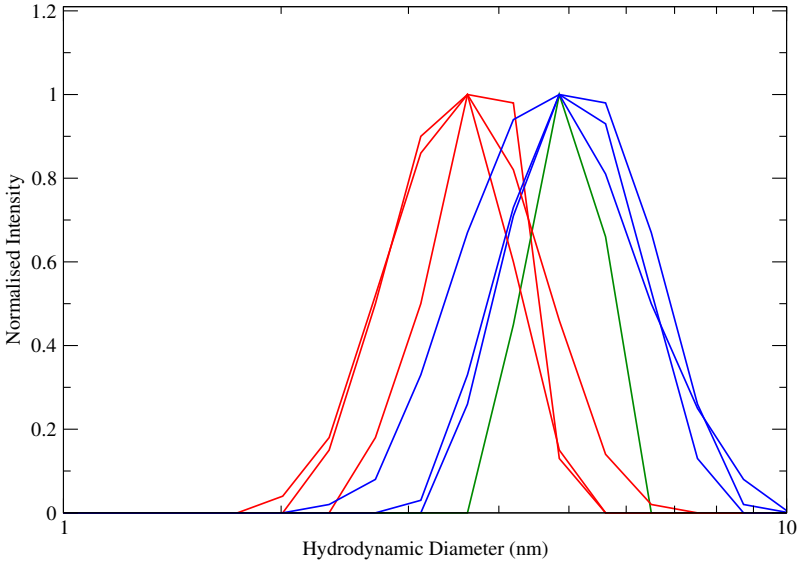


Figure 3.15: DLS data of the disulphide-bonded peptides, taken by EHCB. The red peaks are representative measurements of the disulphide-bonded peptides, 1-6, 2-3 and 4-5, while the blue peaks are the pairs of disulphide-bonded peptides, (1-6, 2-3), (2-3, 4-5) and (4-5, 1-6). The green peak is the mixture of all three disulphide-bonded peptides.

AUC

The measurement and fitting of the AUC data gave no indication that anything other than a single species was present. The molecular weights measured by AUC for the disulphide-bonded peptides and their mixtures (shown in Table 3.4), with the exception of 2-3, were all consistent with the expected molecular weights, within the experimental error given. The measured (within errors) and expected values of 1-6 and 4-5 were consistent with each other, while 2-3, even with a larger error than the other species of the same type, of $\pm$ 2000

Table 3.4: AUC Data: Disulphide-bonded Peptides

Sample	**Expected Molecular Weight**	**Measured Molecular Weight**	$\bar{v}$
6-1	6470	7300 ± 1000	0.746
2-3	7096	9700 ± 2000	0.7482
4-5	6682	7600 ± 1500	0.7458
6-1, 2-3	13566	16000 ± 3000	0.7467
2-3, 4-5	13778	14000 ± 2500	0.7467
4-5, 6-1	13151	16000 ± 3000	0.746
6-1, 2-3, 4-5	20248	19000 ± 4000	0.746

Da, was not consistent; the expected mass was over 2600 Da smaller than the measured mass. The measured mass is consistent with the masses of homotrimers and heterotrimers of single peptides 2 and 3. However, this would only be possible if the peptides were able to reduce their disulphide bonds. The masses measured for 1-6 and 4-5 are consistent with the presence of single disulphide-bonded species, not multiple copies interacting to form larger species. While the mass of 2-3 is not consistent with the mass of a single 2-3 species, it is also inconsistent with a species formed from multiple interacting copies of 2-3. The molecular weights of (6-1, 2-3), (2-3, 4-5) and (4-5, 6-1) are consistent with interactions between a single molecule of each of the two pairings present. The fact that the single disulphide-bonded peptide measurements did not measure species with masses corresponding to multiple copies interacting to form larger structures indicates that, in the mixtures of two disulphide-bonded peptides, it is likely that one of each species is interacting as opposed to two copies of the same molecule. The mass of the whole set of (6-1, 2-3, 4-5) is consistent with a species formed from one copy of each of the disulphide-bonded species present.

3.5 Conclusions

Helicity of the six hub peptides was measured and compared to the peptides when mixed with both designed and non-designed peptide partners. It appears that all of the peptide pairings have a more helical structure than individual contributing peptides, regardless of whether the pair are intended partners. However, the preferential pairing of the de-

signed pairs above all of the other possibilities was strongly supported by the disulphide exchange reaction, which displayed the three desired pairs, and none of the 12 undesired dimer pairings, excepting an excess of 2-3 (and a trace of 4-5), both starting products. Hydrodynamic diameters of the designed hub peptide pairs, measured using DLS, were consistent with blunt ended species. The AUC data was consistent with the interaction of one set of the three disulphide bonded peptides to form a structure. There was no evidence from the AUC data that the disulphide bonded peptides would form larger species than designed. The DLS data of the full set of disulphide bonded peptides also gave a hydrodynamic diameter of the size approximately expected for a hub of the structure suggested in Figure 3.1. This combined information indicates that the hub system is able to form well defined designed structures. The peptides involved are functioning as desired. The main concern is the inconsistent behaviour of the disulphide bonds within the system. Although prepared in the same way, with the same buffer solutions and pH values, the disulphide bonds in one sample of the system were stable whilst in another they rearranged readily. The redox potential of the system needs to be more strongly controlled, or the disulphide bonding in the system replaced by a less sensitive linkage between the three coiled coils of the system.

To further develop the system, a more robust means of connecting the three legs of the Tumbleweed is desirable. It is clear that the disulphide bonds in the system are easily compromised. The synthesis of three long peptides, each with two coiled-coil domains, as opposed to six peptides later linked via disulphide bonding is one option. This could be difficult, due to the general decrease in successful syntheses with increased sequence lengths, yet, as seen in Chapter 5, it is possible. The extra time and effort spent achieving successful long syntheses will likely save effort in the later stages in ensuring that the system does not rearrange and dissociate. Further data on the dimensions of the system, using other techniques, would be advantageous. The synthesis of fluorescently tagged analogues of the hub peptides, and measurement of Förster Resonant Energy Transfer (FRET) between the tags in an analogue hub structure would give dimensional and dynamical information on the system.

Chapter 4

Bar Motor - A Two-footed Design

4.1 Concept

The bar motor is the second motor design discussed in this thesis. It has two legs and feet, in contrast to the Tumbleweed's three. Again, the legs are formed by coiled-coil dimers and it is coiled-coil interactions that will connect the components of the motor. The idea

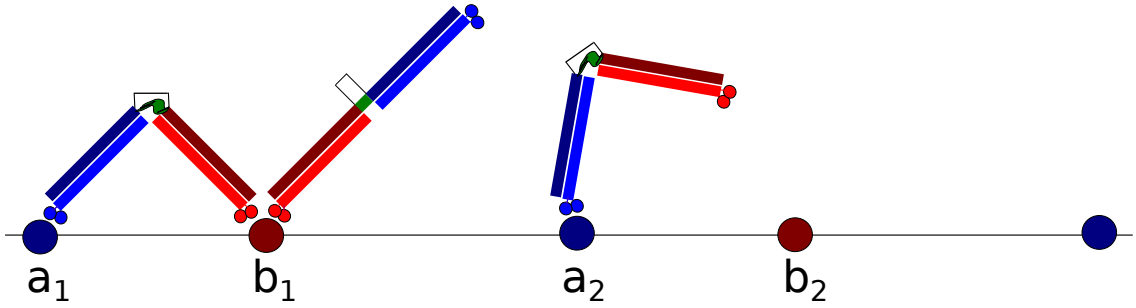


Figure 4.1: The concept of the bar motor. The motor would start in a V conformation bound at sites a_1 and b_1. In order to move the motor there is then a removal of the A ligands from the system, coincident with switching to a straight bar conformation. This is followed by a return of the A ligands, encouraging binding to a_2 (now spatially preferable to a_1). Removal of the B ligands and switching to the V conformation should encourage lifting of the b_1-bound foot, followed by the return of the B ligands to finish the sequence by facilitating binding to the spatially preferred b_2. This leaves the motor in the V conformation bound to a_2 and b_2, one a,b binding site set further along the track. Repetition of this cycle would allow subsequent steps.

is to have a motor with two legs, and a method through which it can switch between two conformations - in this case, a bar-like straight conformation, and a middle-hinged

V-shaped conformation. Figure 4.1 explains the stepping method of the bar motor. In the straight conformation, the separation of the feet is larger than in the V conformation. To encourage progressive motion in the desired direction, the track would have a repeating a,b sequence of binding sites, where (as shown in Figure 4.1) $x_{b_1} - x_{a_1} \neq x_{a_2} - x_{b_1}$. The difference in separation of b_1 with a_1 and a_2 should mean that movement across the shorter distance is unfavourable when the motor is in the bar configuration, and in the V conformation, the closer site should be more favoured due to compatibility between the conformation's feet separation and the track spacing. For these conformations to be possible, the peptide hub needs a central region which is switchable to allow them, and this change needs to be controllable. The initial design idea was to have a long peptide with a flexible central region connecting two coiled-coil sequences which each bind to a shorter partner peptide to form a dimer. These two dimeric coiled coils should give the 'legs' of the hub stable structures. As with the Tumbleweed, the coiled-coil peptides need to be orthogonal so that each motor forms with two different feet. The intention for this design is that when in one conformation, the whole long peptide will be helical and rigid, making a bar-like motor, and in the other, have two rigid legs with a flexible middle, allowing a V shape (along with other conformations between a bar and a hairpin).

This was thought to be possible if the amino acids comprising the central region were chosen to have sufficient inclination for helicity as to allow it, but not be so stable in the helical conformation that it could not be broken by a change in environment. Being surrounded by coiled-coil regions was thought to likely further encourage any natural helicity in the central region. One possible switching mechanism is to have an azobenzene molecule bridging the central region.

Azobenzene is a compound which can be switched between cis and trans conformations using irradiation with different wavelengths of light. Irradiation using wavelengths around 360nm has been used for switching from the trans to cis conformations [196; 197; 198]. The azobenzene isomers have different end-to-end lengths, and it has been used by groups such as that of G. Andrew Woolley as a means of switching peptides between conformations [196; 197; 198; 199]. It isomerizes very quickly, on the scale of picoseconds [199]. As it has been found to be sufficient to disrupt the amino acid chain over which it is attached in some model peptide systems, it may also be able to disrupt the potentially helical central region of the bar motor. If successful, this switching, and coincident removal of ligands, will be sufficient in working against the binding strength of the feet to the track, to force

the motor to remove one foot from the track, and allow subsequent binding to the next, more suitably distanced, binding site.
In summary, the system requirements are:

1. A means of ensuring that the motor forms one unit with two different feet.
2. Two stable switchable motor states.
3. A method of switching from one state to the other in a controlled, repeatable manner.
4. A method of making the steps directional.

Our proposed methods of fulfilling these requirements:

1 a) One component which spans the distance from one foot to the other foot, so that the motor forms one unit - a long peptide sequence with the form (coiled-coil sequence A)-(switchable linking region with two cysteine residues)-(coiled-coil sequence B). In Figure 4.1, this is represented by the dark red-green-dark blue chain.

1 b) Two peptide sequences complementary to sequences A and B (represented by the lighter red and blue rectangles in Figure 4.1), to form dimeric coiled coils. The two dimers need to be orthogonal so each motor has two different feet (one foot will be bound to each of these complementary sequences prior to mixing these with the long peptide to form the whole motor domain).

2&3 A light-switchable azobenzene moiety (the black staple-like component in Figure 4.1) bound to the long peptide via two cysteine residues in its linking region.

4 A directional track as shown in Figure 4.1, and the motor states to have sufficient differences in foot separation to be influenced by the unequally spaced track binding sites.

The design of the track (including the binding sites, the large circular objects in Figure 4.1), and two different DNA binding 'foot' domains (the small circular objects in Figure 4.1), as with those of the Tumbleweed, are outside the scope of this thesis.

4.2 Coarse-grained Simulations

3D coarse-grained (CG) Langevin dynamics simulations were carried out by Martin Zuckermann to complement our work, to check whether the bar motor design would work in

concept. The Langevin equation (originally formulated by Paul Langevin in 1908) used in our simulations describes the Brownian motion of solute particles in a solvent where the solvent molecules are much smaller than the solute particles. It is a stochastic differential equation in which two force terms have been added to Newton's deterministic equations (Newtons second law) to approximate the effects of the fast microscopic degrees of freedom. In this case, the solvent molecules present in the system are not directly simulated and the effect of their collisions with the solute particles is approximated in terms of a viscous drag force on the solute as well as a random force representing the random collisions with the solute particles associated with the thermal motion of the solvent molecules. We use the overdamped version of the Langevin equation, which can be applied to situations such as the motion of our nanomotors where the inertial component of the equation can be disregarded due to the dominance of the viscous drag force (systems which have a low Reynolds number). The equations used for the simulations are included in Appendix G.
The CG model of the bar motor strips away the molecular details of the system, and simply consists of three spheres, representing the two feet, and a hinge (joint that can be varied) between the coiled-coil legs, and two stiff harmonic bonds (of length l) to represent the legs, all connected in a linear conformation (Figure 4.2). The track has repeated a,b binding sites, where

$$x_{b_i} - x_{a_i} = 2l \tag{4.1}$$

$$x_{a_i} - x_{b_{i-1}} < 2l \tag{4.2}$$

and the feet, F_A and F_B, can only bind to a (or b for F_B) when the correct ligand A (or B) is present. As discussed in Section 4.1, the steps of the bar motor are driven by the coincident switching of the motor conformation and ligand concentration change. The conformational switching is represented in these simulations by the changing of the F_A-H-F_B angle, θ_0 (Figure 4.2). A 'pulse' is defined here as a timestep during which a set of parameters are fixed. At the end of each pulse, there is a change in the parameters; here, the ligands present in the system change at the end of every pulse, and the angle θ_0 changes every two pulses. The simulation repeats four 'pulses': (A, 0, ψ), (A, B, ψ), (0, B, ϕ), (A, B, ϕ), where A and B denote the presence of the two ligands, and ψ and ϕ are the values of θ_0. Two simulation conditions have been carried out thus far, both with ψ = 180°, and (a) ϕ = 80° and (b) variable ϕ. For simulation one, $x_{a_i} - x_{b_{i-1}} = 2lsin(\frac{\phi}{2})$. For an l value of 15 nm, this is 19.5 nm. For simulation two, $x_{a_i} - x_{b_{i-1}}$ is set to 23 nm (equating to ∼100°). The simulation is then carried out by solving the overdamped

Langevin equation (Appendix G). In simulation one, the system is directional without pauses (provided there is no external backward force), while in simulation two, the motor could pause due to the flexibility of ϕ allowing the motor to stay at 180° during variable ϕ pulses.

4.2.1 Results and Discussion

Scenario One: Two Fixed Angles

As shown in Figure 4.3, the two feet begin with a separation of 30 nm, indicating the system is in pulse 2, with both ligands present and the motor in the 180° conformation. When pulse 3 begins, foot F_A lifts, as expected, and vibrates around positions approximately centred on the position of bound foot F_B, until pulse 4, and the return of ligand A allows it to rapidly bind at x = 74.5 nm. Once pulse 1 begins, a similar process occurs with foot F_B; it lifts, samples positions within a 30 nm radius, and binds shortly after ligand B is returned to the system in pulse 2 to a binding site at x = 104.5 nm. Each pulse is 2.88 ms, and after 4 pulses, 11.52 ms, each foot has taken a step, moving the whole motor ahead by 49.5 nm (one 30 nm and one 19.5 nm step). With two fixed angles, the motor is processive, does not backstep, does not stall, and appears to bind rapidly after the required ligand is reintroduced to the system.

Scenario Two: One Fixed Angle, One Flexible Angle

Scenario two is a more realistic scenario. Even in the design stages, prior to experiment, it was hoped that rigid and flexible states could be created. It was always expected that the flexible state would undergo a range of different angles. This scenario represents that: half of the simulation pulses have a fixed angle of 180°, while the other half allow the angle to be varied. The results of this are shown in Figure 4.4. For this simulation, a spacing corresponding to a 100° angle was chosen for $x_{a_i} - x_{b_{i-1}}$. It is clear from both axes of Figure 4.4 that there are three missed (pairs of) steps. After the first step of each foot, foot F_B (blue) lifts, but upon the return of ligand B, replaces its foot in the same place as it was previously bound (x = 78 nm). This means that foot F_A has to remain at its (x = 108 nm) position during its next pulse changes, instead of stepping forward. In this case, foot F_B misses two possible stepping opportunities, before stepping on the third attempt. Another single step per foot is missed at t ∼ 1.75 s. In total, after 2.5 s, and 13 four-pulse cycles, 10 cycles led to successful stepping. Each foot progressed by 530 nm, of a possible

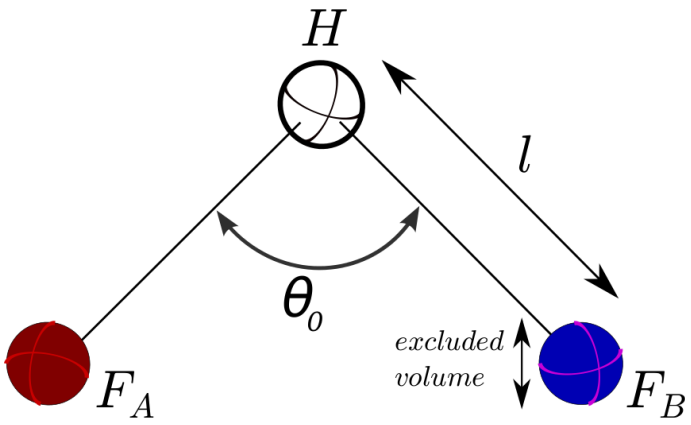


Figure 4.2: The coarse-grained model has three spheres to represent the two feet, and the azobenzene/hinge region. The coiled coils are represented by two stiff bonds.

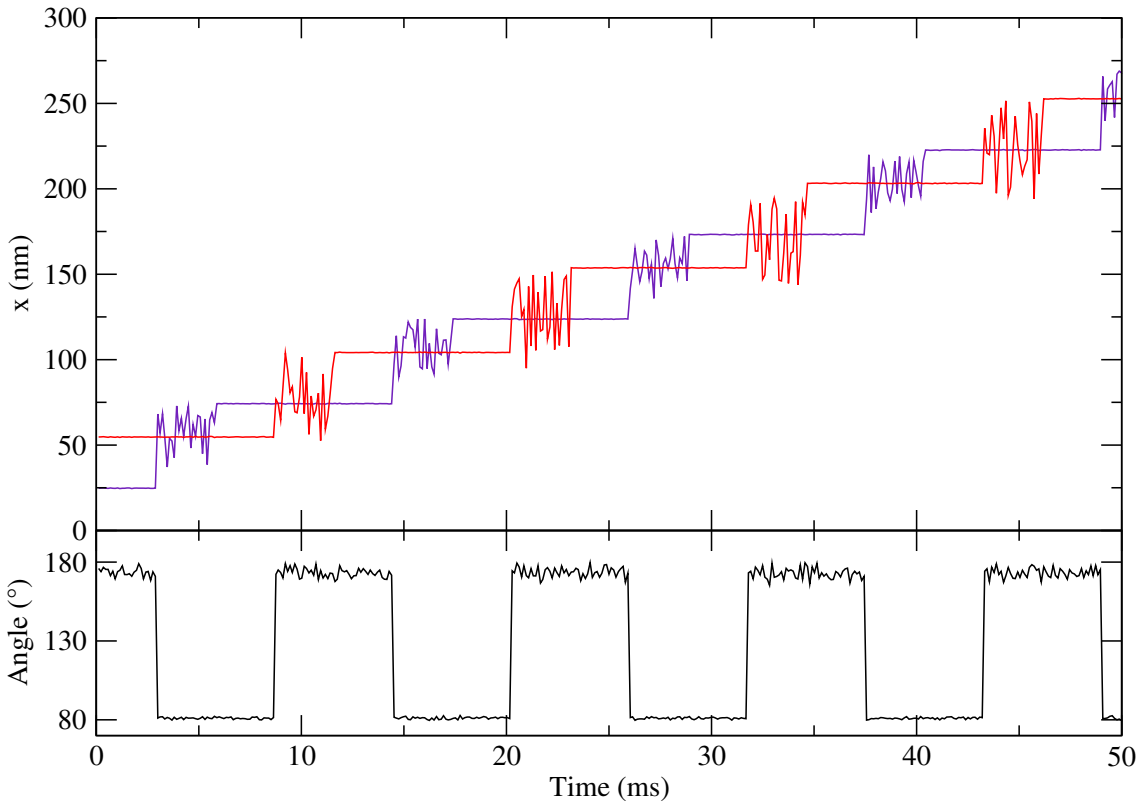


Figure 4.3: Scenario One: stepping of a motor with θ_0 switching between 180° and 80°. The leg length is 15 nm, and the track separations are 30 nm (to fit well with a motor where $\theta_0 = 180°$) and 19.5 nm (80°). Pulses are 2.88 ms long, and hence there is an angle switch every 5.76 ms. The two lines (purple, red) in the top graph show the x-coordinates of the two feet, while the lower graph shows the angle θ_0, as a function of time.

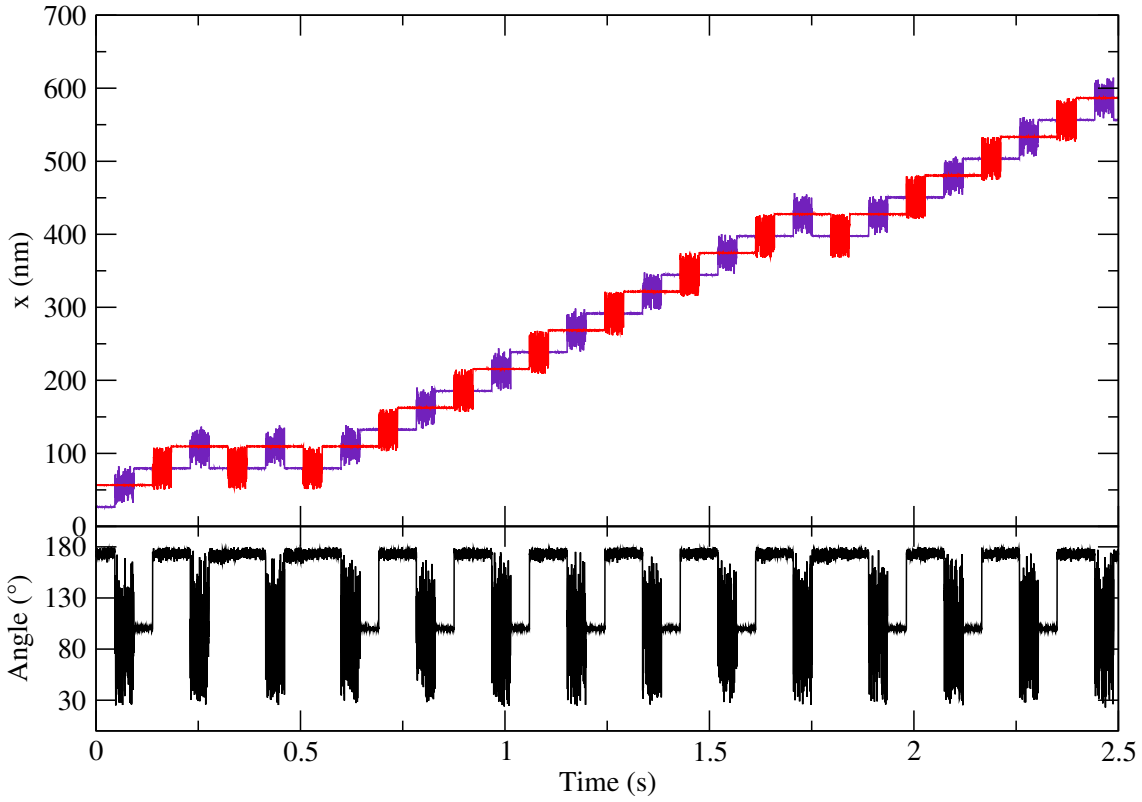


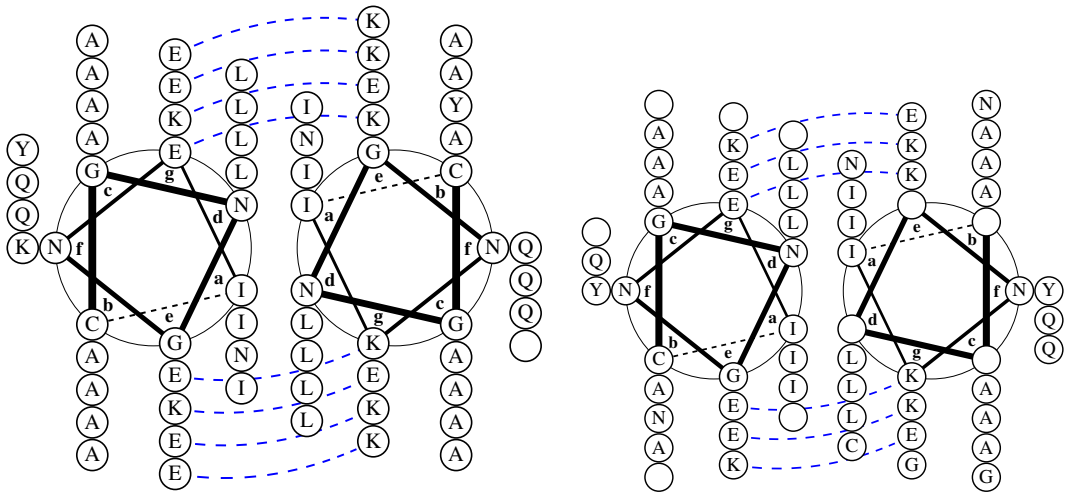
Figure 4.4: Scenario Two: stepping of a motor with θ_0 at 180°, but with the binding energy potential value switched, between fixed and variable θ_0. The leg length is 15 nm, and the track separations are 30 nm (to fit well with a motor where $\theta_0 = 180°$) and 23 nm (equating to 100°). Each pulse is 46 ms long. Again, the two lines (purple, red) in the top graph show the x-coordinates of the two feet, while the lower graph shows the angle θ_0, as a function of time.

689 nm, giving an efficiency of 77%. Whilst this is clearly less efficient that a fixed angle motor, this indicates that a motor system much closer to that which we will be able to create is still able to make progress.

4.3 System Design

The first attempt at experimentally testing the concept was the design and synthesis of a 59-amino acid peptide which interacts with peptides 1 and 6 (two of the Tumbleweed peptides). This long peptide will be henceforth known as ϵ2, due to its being based on the sequences of peptide 2 (previously discussed), and a sequence known as ϵ. ϵ is an alternative peptide to 5 as a partner for peptide 6, and forms a parallel coiled coil with peptide 6 instead of the 5,6 antiparallel coiled coil. ϵ,6 was originally one of the pairs for the triangular version of the Tumbleweed hub (Figure 1.19c). The helical wheel diagrams for 1,2 and ϵ,6 are shown in Figure 4.5. In addition to its forming coiled coils with partners, a desired characteristic of the long peptide is that the central region (replacing those residues intended to form a flexible linker in the Tumbleweed peptides ϵ and 6) forms neither a highly stable helix nor is completely unfolded, so that it can be switched from rigid to flexible. This central region therefore needed to be adapted from the original sequences.

Figure 4.6 shows the steps of designing a suitable long peptide sequence for the first incarnation of the bar motor. The sequences of peptides ϵ and 2 were put together, and the linking regions shifted to the centre (steps 1 and 2). In the triangle design, ϵ was disulphide-bound to another peptide at the N-terminus, hence the linking region being at the N-terminus, but here it will be covalently linked to 2 at the C-terminus. It is at this point that the requirements for the azobenzene moiety had to be taken into consideration. Figure 4.7 shows the azobenzene linker synthesised by Asahi Cano-Marques (ACM). The trans isomer is the more energetically favoured, and has the larger end-to-end separation. It can be bound to a peptide via cysteine residues. With an (i, i+7) spacing between the involved cysteine residues, the azobenzene's trans state will cause a separation between the linked residues too large to allow helical structure between them [197] (Figure 4.8). When the azobenzene is switched to the cis isomer, the residues should be brought together close enough for helicity to be possible. Helicity is compatible with an (i, i+7) separation of cysteine residues and an azobenzene cis isomer, or an (i, i+11) separation and azobenzene trans isomer [197; 196]. Switching of the azobenzene is done using different wavelengths



(a) 1,2 (parallel) (b) ϵ,6 (parallel)

Figure 4.5: Helical wheel diagrams (made using DrawCoil [161]) showing the interactions between pairs of peptides in the Bar motor. Dashed lines show salt bridges between e and g residues. The blank circles are residues absent from the peptide, but which have a corresponding residue in the partner (for example, peptide 1 has one more residue than peptide 2, ending with an f residue, while peptide 2 ends with an e residue (see Table 1.1).

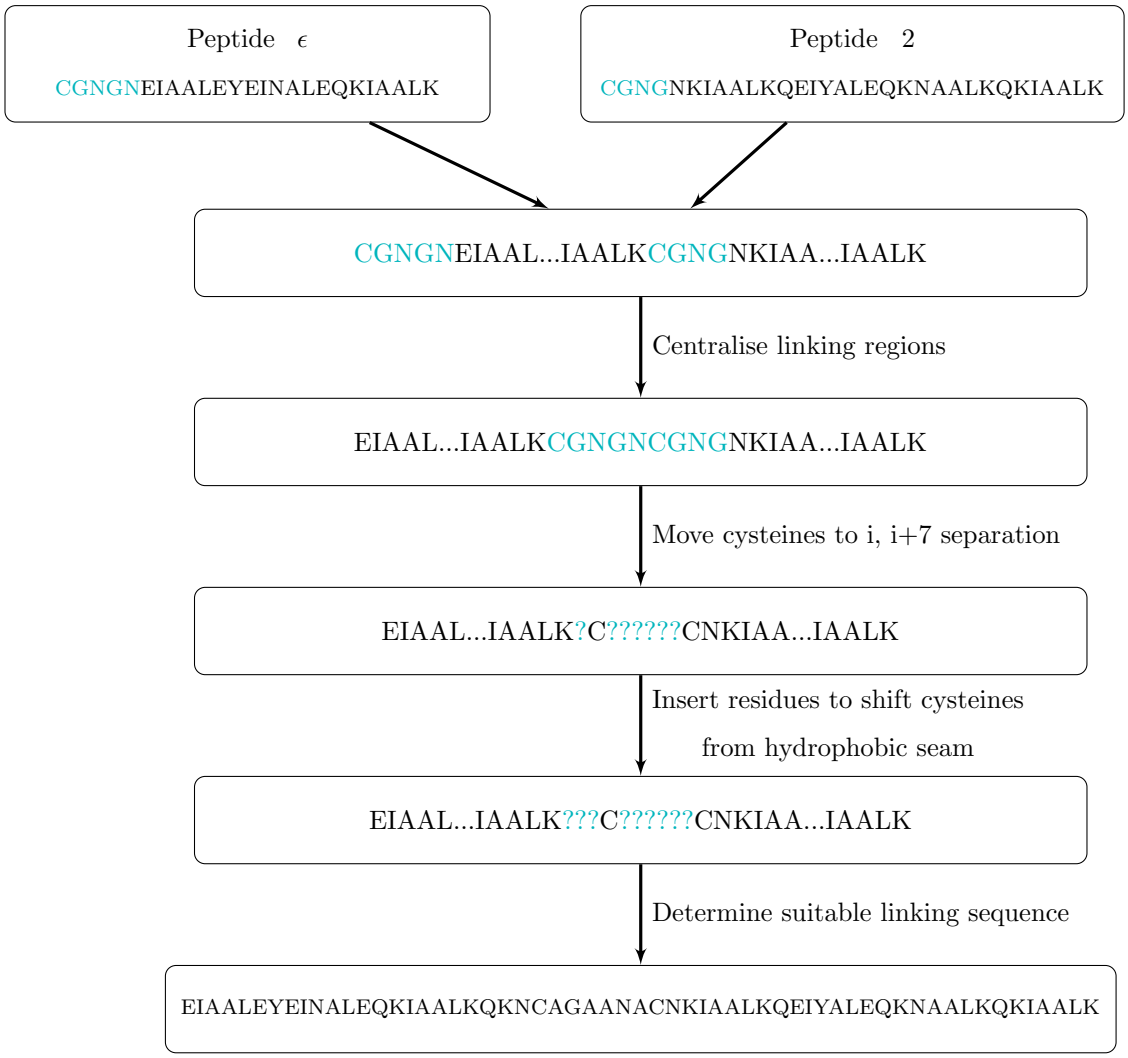


Figure 4.6: Steps of designing the bar motor sequence.

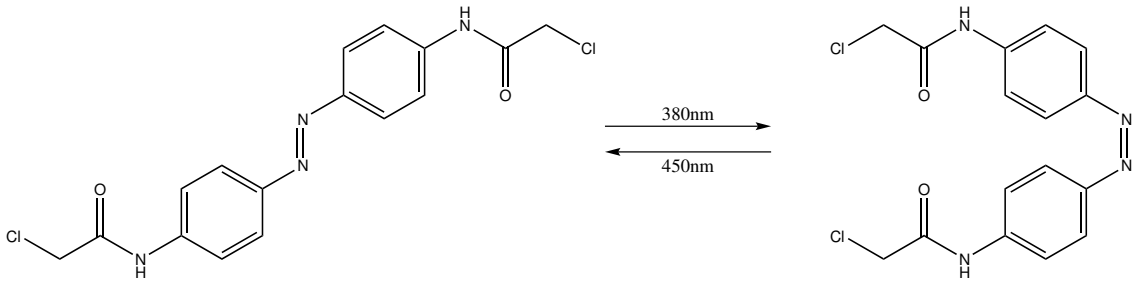


Figure 4.7: Azobenzene linker synthesised by ACM. The azobenzene molecule switches isomer when photoexcited by 380 nm light. The trans (left) and cis (right) isomers are shown. As implied by the figure, the isomers have different end-to-end separations, which is the property we wish to exploit.

of irradiation, and the success of it dependent on whether this switching is strong enough to overcome the interactions causing the helicity. In this case, we placed the two cysteines

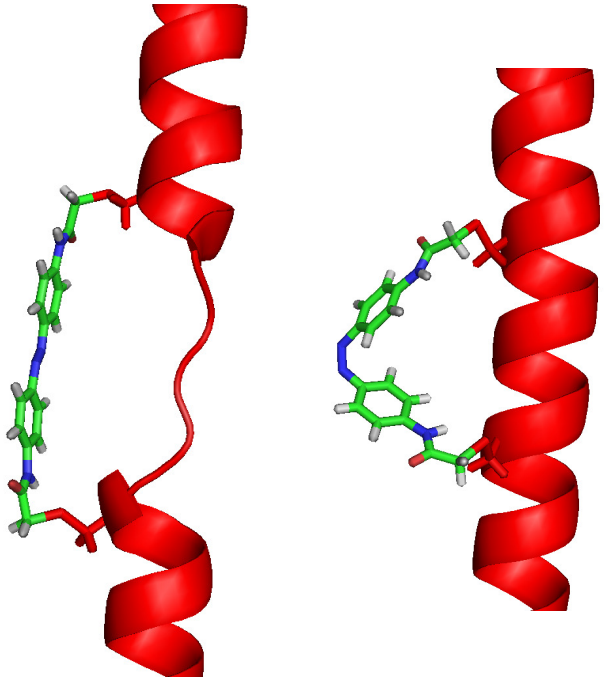


Figure 4.8: For a 7 residue separation between the azobenzene ends, the azobenzene in the trans state should provide an end-to-end separation sufficiently long as to distort the contained residues' helicity. When the azobenzene is switched to the cis state, the extremities of the azobenzene are brought together, allowing the surrounded residues to be close enough to form a helix.

7 residues apart, and all of the other central region amino acids were set to undefined (Step 3 of Figure 4.6). We added two residues before the first cysteine to shift the cys-

teines to b positions, towards the outside of the coiled coil (Step 4); in a or d positions they are more likely to interfere with coiled-coil formation in an unintended manner, by disrupting the formation of a hydrophobic seam. Various alanine, asparagine and glycine combinations were tested as possible six residue linking region sequences using AGADIR [200; 201; 202; 203; 204].

AGADIR is an algorithm for calculating the helicity of peptides in solution, using helix-coil transition theory. AGADIR uses early helix-coil transition models (describing the transition from α-helix to random coil) by Zimm & Bragg [205] and Lifson & Roig [206] as a basis, which contain nucleation and elongation/propagation parameters - whether a residue was likely to initiate a helical region, and/or continue an existing one, along with more recent information. It includes hydrogen bonding, capping interactions, the intrinsic helical tendencies of given residues, charge-dipole interactions, interactions between side chains, and some local motifs which stabilise helices. It includes the effect of temperature, ionic strength, and pH. It gives a residue-by-residue helicity content, which is the data we extracted and used here to influence our choice of central sequence.

Various linking regions were inserted into the sequence and assessed using AGADIR. The results are shown in Figure 4.9. The helicities found by the program are low in general for the whole peptide, but this is unsurprising, as the peptides are designed to be helical when forming coiled coils with a partner, not to necessarily be naturally helical alone and AGADIR does not have this coiled-coil information. Setting aside this generally low helicity percentage, the AGADIR results show that the linking region (independent of the sequence within the test set used) has a lower helicity than the two would-be coiled-coil regions, which we were aiming for. Of the tested sequences, an all-alanine linker was found to be the most helical of the test sequences, as one might expect from previous studies (discussed in Section 1.4.2). It is then difficult to determine which of the other sequences are least helical, due to the small differences between them. It is clear that the least helical sequences generally had less alanine, with the least helical, GAGGAG, having no sequential alanine residues. Asparagine residues seemed to be less disruptive to helicity than glycine, but sequence position was important (likely due to whether a residue broke an alanine chain or not). The sequence chosen for the centre of $\epsilon 2$ was somewhere between the two extremes, to hopefully allow switching: AGAANA.

Putting both peptide sequences together with linking regions in the centre meant that the two coiled coils were out of register. If labelling the residues a-g straight through (starting

at g), the second half had hydrophobic residues at e and a instead of a and d positions. This still leaves the hydrophobic residues on the right side of the helix. By shifting the positions of the cysteine residues through the insertion of additional residues, we have put in a stutter [207] which may result in a straight coiled coil, which would be desirable for a bar motor. The sequence chosen was EIAALEYEINALEQKIAALKQKNCAGAANAC-NKIAALKQEIYALEQKNAALKQKIAALK.

As discussed in Section 4.1, for the bar motor to be successful, conformational differences must exist between the cis and trans azobenzene-containing peptide systems. The angles open to the two conformations need to be sufficiently different so as to make the motor processive and resilient to backstepping. The CG simulations showed that in a rigid and flexible switching system, though there were some missteps, there was still progression, and hence feasibility in the design. Our next step, in addition to carrying out experiments, was to gain some MD data to investigate whether the system as designed would be able to fulfill this criteria.

4.4 Molecular Dynamics Simulations

To study the angles that the molecules are likely to sample, we are interested in the actions of individual sets of interacting peptides, and hence atomistic simulations are required. Computational power limits the system size and timescale for which a system can be simulated, due to the number of atoms and bonds present, which are all modelled in these simulations. A less complex system (or simplified version of a complex system) can be simulated for a longer timescale in the same amount of real time, which is why in some simulations, such as the proof of concept simulations (Section 4.2), coarse-graining is useful. As we are interested in both the motion possible, and the timescales of this motion, Molecular Dynamics (MD), with its simulation of a molecule's time evolution, is suited to the task.

First demonstrated in 1957 [208], MD is frequently used for systems such as ours, due to this ability to provide dynamical information on an atomic scale. It is completely classical. One copy of the molecule (or molecules) to be simulated is put into a periodic box (surrounded by identical boxes of identical molecules), the box is filled with water (in our case), and forces (described by a force field) act on the molecule. The simulation progresses in small time steps, where Newton's equations are solved to find the force on each particle, and hence their velocities and positions at each time step. The simulation

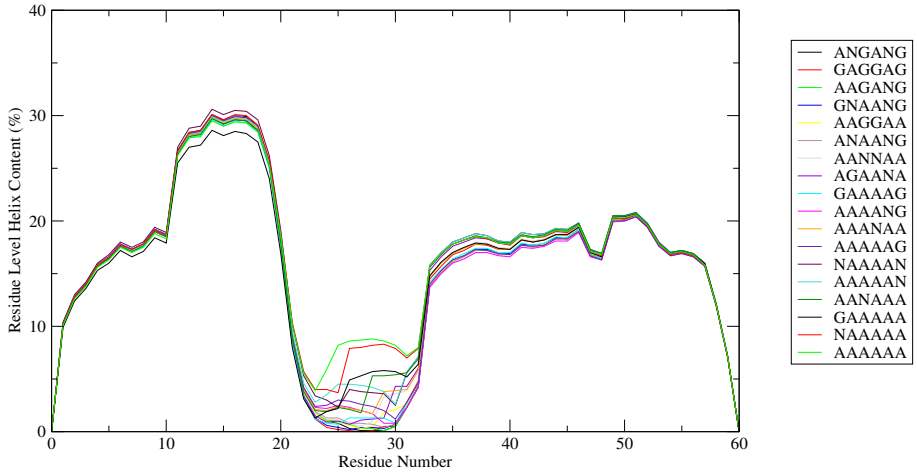


Figure 4.9: Results of test sequences in AGADIR. Each sequence was EIAALEYEINALEQKIAALKQKNC******CNKIAALKQEIYALEQKNAALKQKIAALK where ****** is the sequence shown in the legend. The coiled-coil regions are shown to be fairly constant in helicity level with changing central sequence.

outputs the trajectory of the molecule as a function of time.

GROMACS [209; 210; 211; 212; 213] is a popular MD package due to its user friendliness, speed, availability, compatibility with a range of force fields and water models, and its lacking the requirement of specialist computing equipment; as the tagline states, it is 'Fast. Free. Flexible.' [211]. GROMACS requires the initial coordinates of the atoms in the molecule(s) to be simulated; we used models constructed by Richard B. Sessions (RBS) using Insight II (Accelrys Inc, 2005) from the sequences we provided (LSRS and EHCB). Models of the three free peptides, ϵ2, 1 and 6, and the three peptides with an azobenzene moiety on ϵ2, in each of the cis and trans isomer conformations, were produced (Figure 4.10). We have made the assumption, based on the sequences and previous Tumbleweed data, that the (1,2) and (ϵ,6) domains of the system will be helical, and have assumed the central region to be helical unless disallowed by the dimensions of the azobenzene molecule (the model with azobenzene in the trans conformation). As is seen in the results, the simulations then allow this region to explore alternative configurations. Additionally, one provides GROMACS with a model for the water surrounding and interacting with the simulation molecule(s); TIP3P, an explicit water model, is used in these simulations. Many force fields, or potential energy functions, as they are also known, can be used with GROMACS, providing the coefficients needed for force calculations of various molecular interactions. These interactions include those between non-bonded atoms, such as the

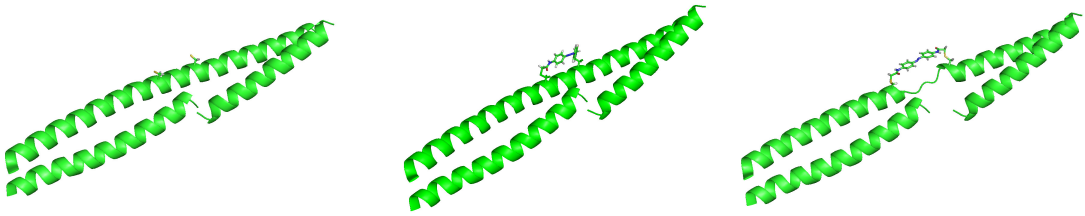


Figure 4.10: Models used for simulations. Peptides 1 and 6 with (a) $\epsilon 2$, (b) $\epsilon 2$ with cis azobenzene linker across Cys24 and Cys 31, and (c) $\epsilon 2$ with trans azobenzene linker.

Coulomb interaction, repulsion (due to electron cloud overlap) and dispersion (attraction due to instantaneous multipoles), and bonded interactions: bond angles, lengths, torsions and dihedrals. We use the Amber ff99SB-ILDN force field. The TIP3P water model was chosen for the simulations as the Amber force field was developed using TIP3P [214], and the combination of the two has been used extensively over the last 20 years. The Amber ff99SB-ILDN - TIP3P combination has been found to give a reasonable description of peptide/protein systems when compared with experimental data [215; 216; 217]. Additionally, periodic boundary conditions [218] are used in the simulations to avoid the difficulty in modelling the differing behaviour of molecules at surfaces: a box containing one copy of the molecules to be simulated is surrounded by identical boxes of identical molecules, and if a molecule leaves one box from one side, it is replaced by the identically behaving molecule from the box on the opposite side. The minimum image convention is used so that a particle can't see the copies of itself, and only one copy of any neighbouring particles also present in each box; to do so, a cut off of half (or less) of the shortest box dimension [219; 220] is applied for these short range non-bonded interactions. A sensible choice of box type (in this case, dodecahedral) minimises the proportion of the simulation that is just between water molecules (as opposed to between atoms of the peptides, or peptide-water interactions), and reduces computing time.

4.5 Initial MD Study

Upon construction of the peptide models, RBS ran a preliminary 100 ns simulation of each of the three systems which were analysed by myself. Visual observation of the simulations as movies gave an impression of the behaviour. Of course, repeat simulations are needed to give more information, and determine which motion is most common, but these give an

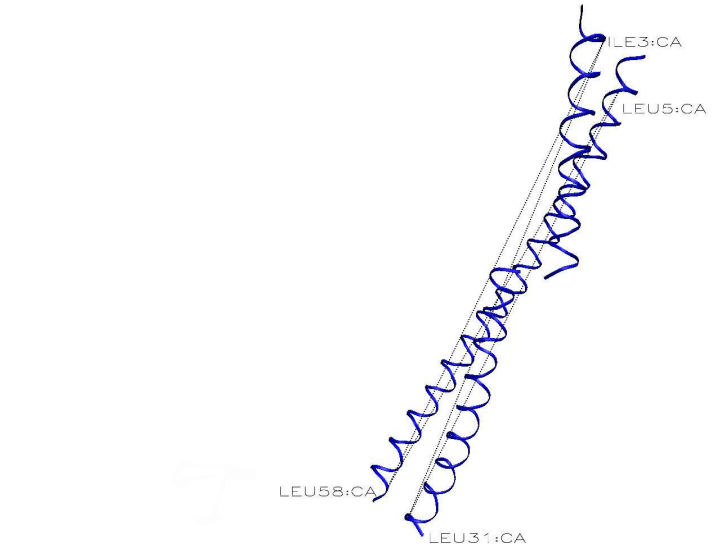
indication of the behaviour possible. For all three systems, the coiled coils remained intact throughout the simulations, indicating that the models were not too unstable to function, and worth further consideration. In order to quantify the useful motion of the peptide system, the varying end-to-end separation of the peptides, and angle made between the two coiled coils at the central hinge, were extracted, as it is a difference in these which is necessary for the system to function.

4.5.1 Results

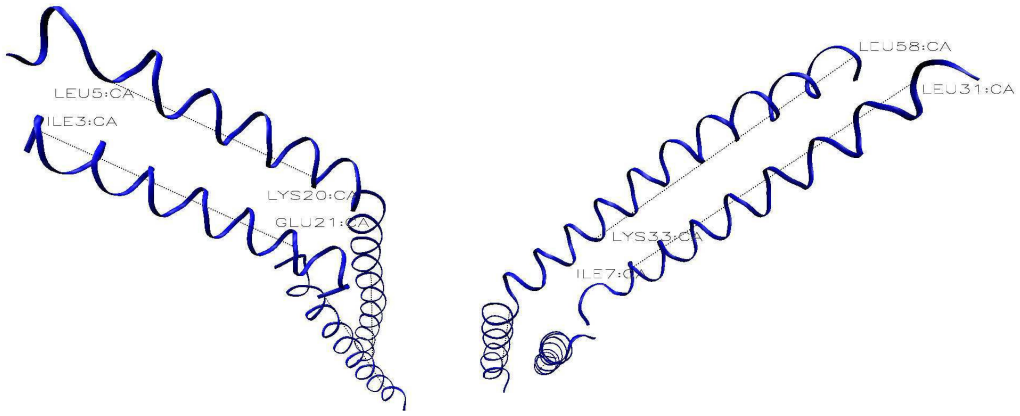
End-to-end Separation Measurements

The separation between the α-carbon atoms of two residues is easily extractable using VMD (Visual Molecular Dynamics) [221], a molecular graphics program. Some of the extremely-placed residues were seen to uncoil, and not represent the movement of their coiled coil as a whole. In order to ensure that the data extracted was representative, four separation measurements were taken, instead of one, between the four pairs of opposite peptide termini (Figure 4.11a). The α-carbons of the extremity-positioned residues Leu5 and Leu58 of $\epsilon 2$, Leu31 of Peptide 1 and Ile3 of Peptide 6 (unless stated otherwise) were used, as they were at extreme positions, but did not frequently uncoil. To confirm that the chosen residues were suitably stable relative to the rest of their coiled coil, the measurements shown in Figure 4.11b were taken (a sample is shown in Appendix H.1). As each of these separations are between an outer residue and inner residue of a single peptide, within a single coiled-coil region, they should be close to constant in a stable coiled coil. If any of these measurements changed significantly over the course of the simulation, i.e. there was structure in the time evolution instead of a noise-like spectrum, due to a particular residue, the residue was usually also seen visibly to leave the stable coiled-coil region for some of the simulation time. Frequently, it was proven that it was the internal residue which was causing the variations, when the outer residue was compared with an alternative inner residue. In cases where the outer residue was the source, those separations were treated with more caution than the other three measurements of the system.

Upon inspection of all of the measurements for each system, it became clear that there was no significant difference in the shape of the four measurements, despite any occasional erratic motions, and so an average of the four was taken and plotted here, to more clearly display the system's behaviour.

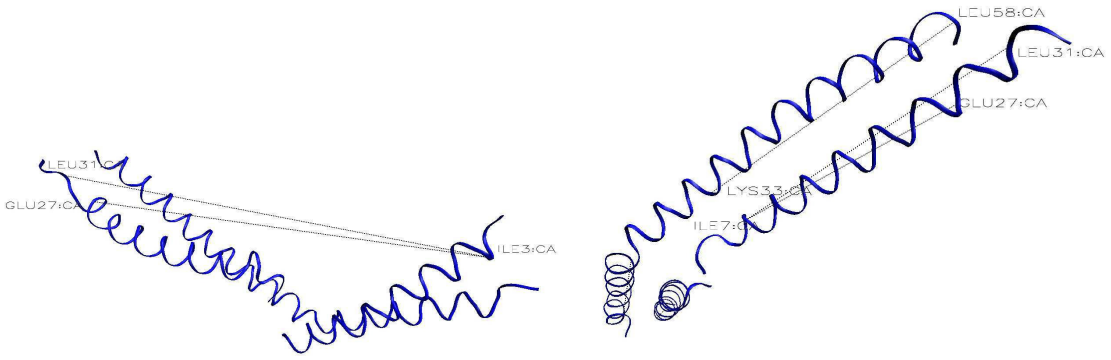


(a)



(b)

Figure 4.11: Four values of the end-to-end separation were found for each system to avoid any anomalous measurements due to a single residue, using the combinations of $\epsilon 2$'s Leu5 and Leu58, peptide 1's Leu31 and peptide 6's Ile3 shown in (a). To confirm that the chosen residues remained part of coiled-coil structure and didn't uncoil, the evolution of the separations (between each of the residues in (a) and a stable internal residue of the same coiled coil and peptide chain) shown in (b) was checked, to ensure that the evolution of the separations in (a) represented the motion of the whole system, and not the motion of a short uncoiled region at the peptide extremities. For the cis linked system, Glu27 was used as well as Leu31 in the measurements, as Leu31 formed part of the region shown uncoiling in (c). As with the other residues, the constancy of Glu27 within its coiled coil was checked (d).



(c) (d)

Figure 4.11: For the cis linked system, Glu27 was used as well as Leu31 in the measurements, as Leu31 formed part of the region shown uncoiling in (c). As with the other residues, the constancy of Glu27 within its coiled coil was checked (d).

Central Angle Measurements

Angles between three α-carbons are also easily extractable from VMD, but choosing residues for this was the more difficult task, as the 'centre' of the system is not a well defined point. For each of the systems, the evolution of the angle formed by Leu5, Lys52, and one of Lys20, Asn23 and Ala27 on peptide $\epsilon 2$ was extracted (Figure 4.12a), and for the azobenzene-free and cis azobenzene systems, this method gave a good graphical representation of the motion of the system. For the trans azobenzene system simulation, the central region was far more flexible, frequently with more than one turning point, and this made the calculation of a representative angle more difficult. The inner ends of the coiled coils did not stay together in the centre of the system, due to the increased flexibility of the central region. This did not affect the method of end-to-end separation measurements, but meant that choosing a central residue for the angular measurements was problematic and often an produced an uninformative measurement. Placing three points on the 'T' shaped structure in Figure 4.12b to approximate the angle calculated by this method demonstrates its failure for the trans azobenzene simulation. To produce a more insightful angular measurement, I produced vectors representing the axes of the two coiled coils, and found the corresponding angle if the vectors had a common origin (Appendix H.2). For this, a pair of a or d residues near each end of each coiled coil (but well within the stable coiled-coil structure) was used (Figure 4.12c), as they were thought to be stable residues to produce accurate vectors. The separations shown in Figure 4.12c were confirmed to be stable, and

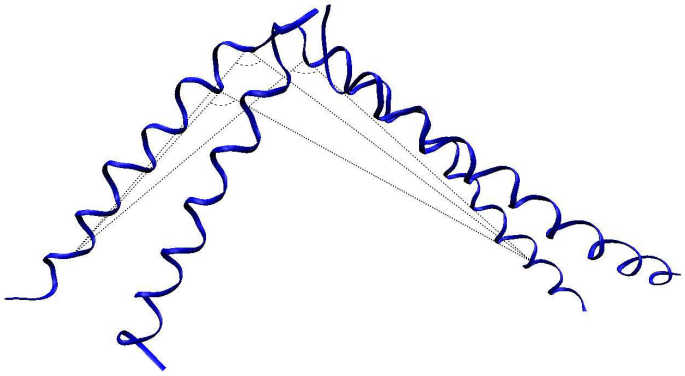
then midpoints between each pair of residues (Leu5 and Leu6 is one pair, for example, Figure 4.12c) were used to create two vectors representing the coiled-coil axes, and calculate a subsequent angle. For the systems where the molecules have hinge-like behaviour, this method should be equivalent to the three-point method, and this was confirmed for the first azobenzene-free and cis azobenzene simulations. End-to-end separations for the trans system using the residues in the vector method were also calculated and found to be consistent with the patterns seen using the extreme residues. The residues used (a midpoint between ϵ2's Ile55, and p1's Ile28, as opposed to the α-carbons of both Leu58 and Leu31, and a midpoint between ϵ2's Leu5 and p6's Leu6 instead of the α-carbons of both Leu5 and Ile3) were generally further from the ends than the VMD-calculated end-to-end separation residues, and hence the separations appear generally lower. The original separations therefore give values closer to the actual separations of the ends of the peptide systems during the simulations, being from residues closer to the extremities, and continued to be used. Of course, even these vector measurements do not display the lack of central origin to the observer, and hence must be used in conjunction with the movies, and with an awareness of their limitations.

4.5.2 Discussion

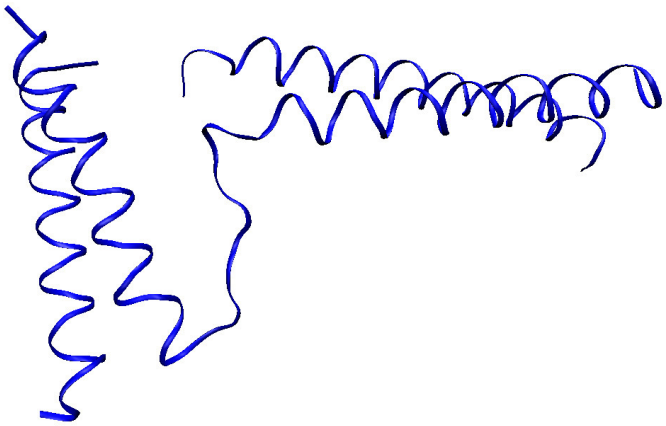
The ϵ2 starting conformations for the cis and azobenzene-free systems have no clear break in helicity, as breaking the helicity of the peptide was not necessary to attach a cis azobenzene to the desired residues. For the trans isomer, the dimensions of the azobenzene molecule did not allow it to be bound to the correct peptide residues whilst the ϵ2 peptide was fully helical, and hence the secondary structure had to be modified to accommodate it (Figure 4.10).

Azobenzene-free System

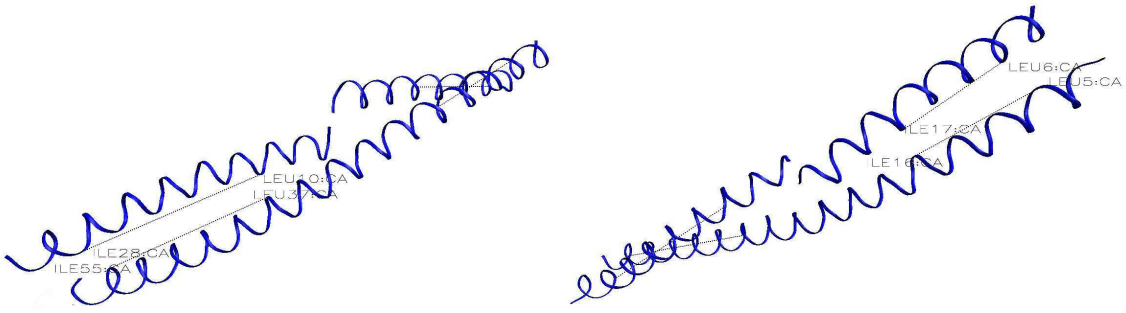
The 100 ns simulation indicated that the unmodified three-peptide system could exhibit a considerable angular range, from around 180° (outstretched), to near-90° angles (Figures 4.13 and 4.16). From a movie of the simulation, one can see that the system undergoes rapid fluctuations, as expected from a dynamic molecule. It does not progress in one direction from 180° to smaller angles without fluctuations to larger angles, but there is only one eventual movement to a small angle; the system does not move back and forth between the outstretched state and more compact state within the simulation time.



(a)



(b)



(c)

Figure 4.12: (a) The three-point angle method is simple using VMD. (b) An example from the trans azobenzene simulation where the three-point method fails - the angle calculated, between the residues labelled in (a), does not follow the shape of the system. (c) Angles formed between two vectors were calculated as an alternative to the three-point angle method.

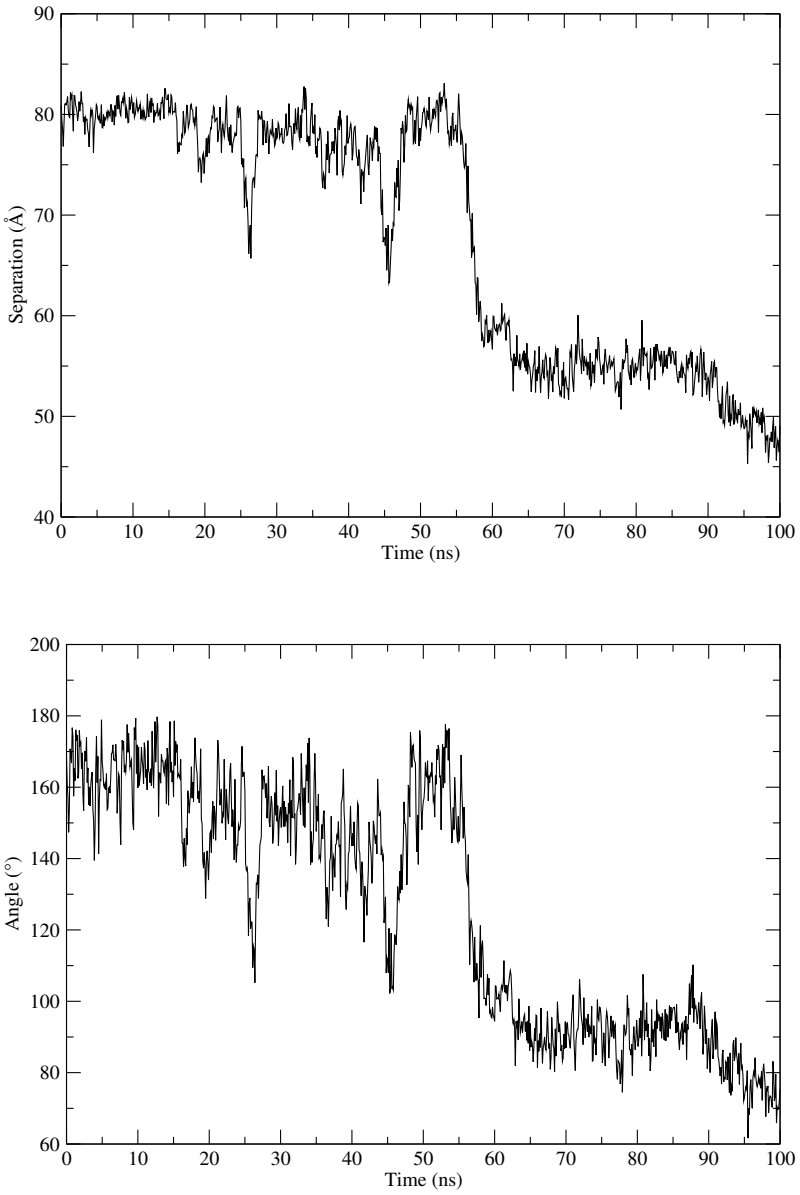


Figure 4.13: Evolution of the end-to-end separation and hinge angle of the azobenzene-free bar system over a 100 ns simulation.

The graphs of both end-to-end separation and angular variation (Figure 4.13) display two significant regions, almost as if the system is occupying two significant states, with fluctuations around them. Both graphs have the same features, as most of the movement affecting the angle and the separation is in the same plane. The system's near-straight starting conformation corresponds to a separation between the bar ends of around 8 nm (with a 160° central angle). The system fluctuates around this state for around 60 ns, after which it switches rapidly to a conformation giving an end-to-end separation of around 5.5

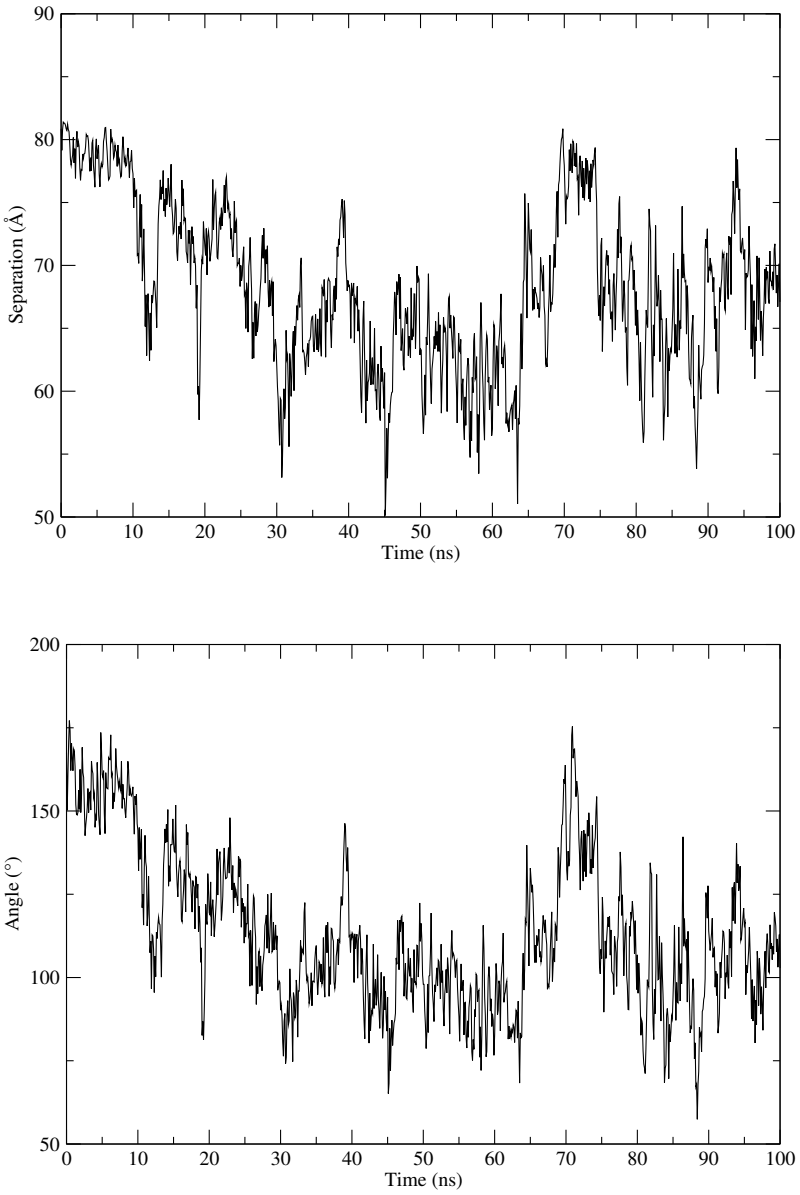


Figure 4.14: Evolution of the end-to-end separation and hinge angle of the cis azobenzene-containing bar system over a 100 ns simulation.

nm (and an 80° angle). The N-terminal residues of peptide 6 uncoil so that the inner residues can overlap - if they remained helical there would likely be greater restrictions to the angle possible in the cases where the partner peptides are held closer as a result of the bending (as is the case in most of these simulations where bending is seen). It fluctuates around this region until near the end of the simulation, where it looks as if the system may be undergoing a more gradual change to an even smaller end-to-end separation/angle. This system appears to undergo steady motion, with no erratic changes, but as it lacks

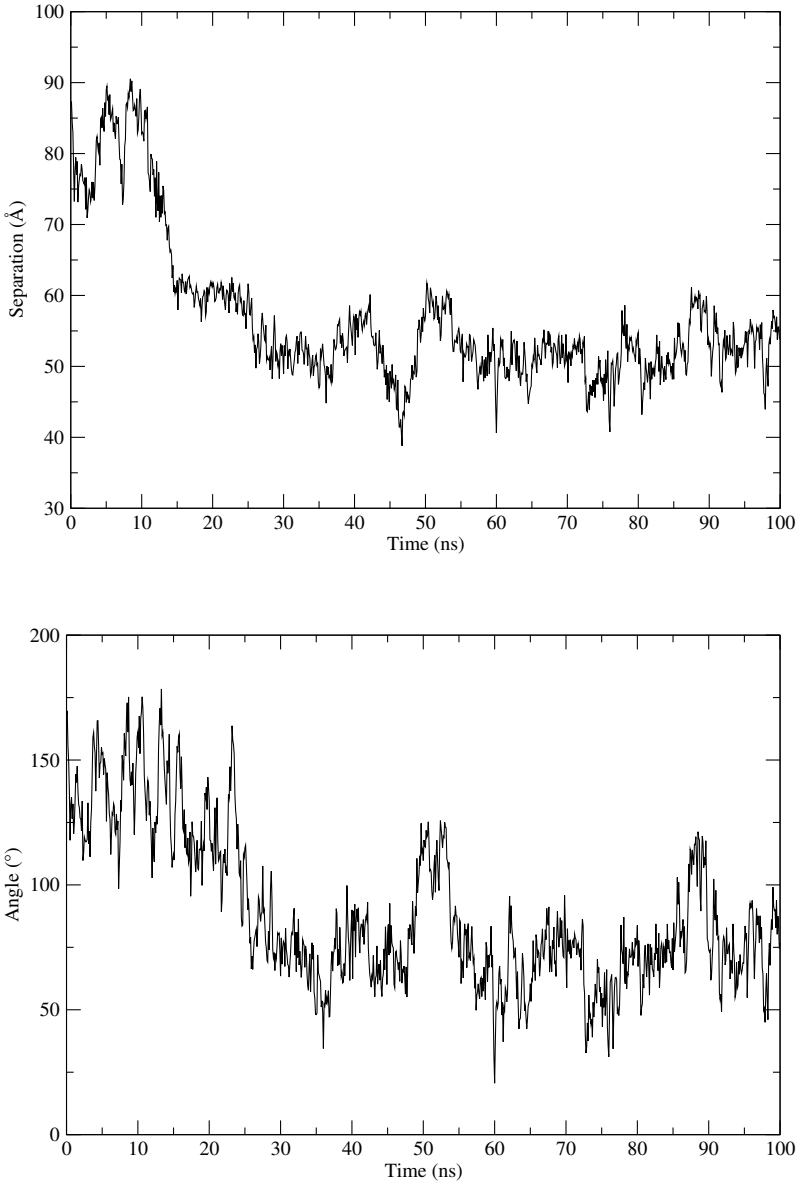
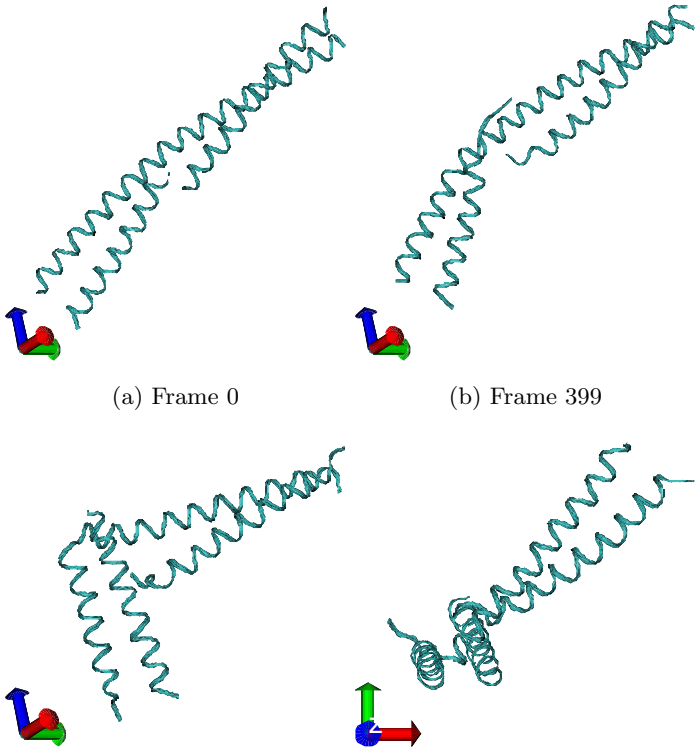


Figure 4.15: Evolution of the end-to-end separation and hinge angle of the trans azobenzene-containing bar system over a 100 ns simulation.

a means of being switched between two or more states, is not useful unmodified as a bar motor hub. The simulation indicates that there is no cause due to the peptides in the system for the angles required to be unattainable.

Cis-azobenzene System

Figure 4.17 shows some of the frames from the cis azobenzene simulation, while Figure 4.14 shows the end-to-end separation and angle evolution of the system. For the cis azobenzene system, Glu27 was used in addition to Leu31 due to Leu31's uncoiling during

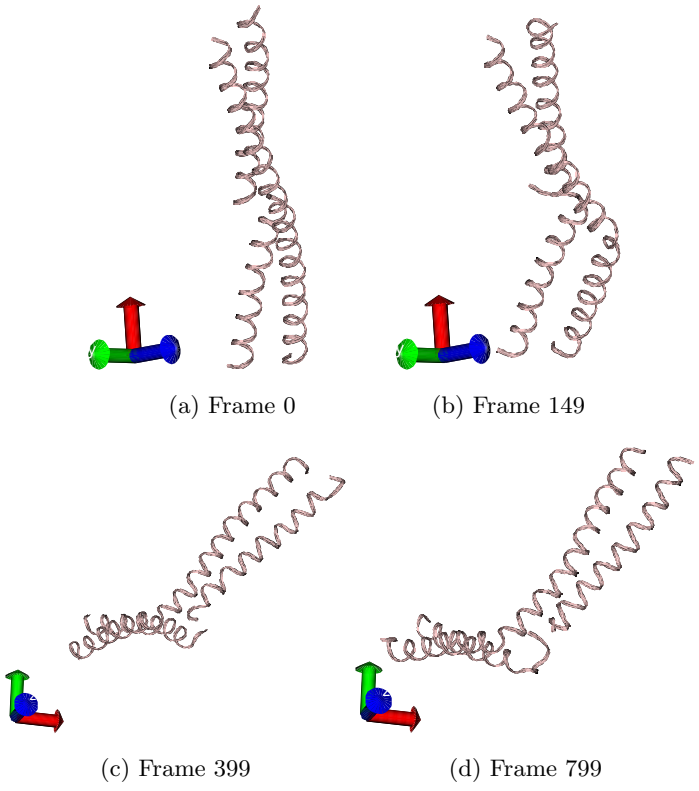


(a) Frame 0 (b) Frame 399

(c) Frame 599 (d) Frame 599, view 2

Figure 4.16: Stills from the initial 100 ns simulation of the azobenzene-free bar system.

the simulation (Figure 4.11c, d). However, it was found that although Leu31 visibly uncoiled during the simulation, this did not significantly alter its separation from its coiled coil (Figure H.1), and end-to-end separations using Leu31 still exhibited the same patterns as the other measurements, including those of its substitute, Glu27 (Figure 4.11c). The Leu31 measurements were therefore used in end-to-end separation averages. Visually, the cis system seems to have a similar separation range open to it as the azobenzene-free system. However, the motion does not have the two distinct regions of the azobenzene-free system (graphically more obvious by comparing Figures 4.13 and 4.14). As this is only based on a single 100 ns simulation of each it is likely not to be a characteristic feature, but future simulations will confirm this. As is shown by the figures, the system first enters a sub-90° conformation earlier in the simulation than in the azobenzene-free simulation. The angular variation patterns again mimic the patterns seen in the end-to-end separation evolution, with a slightly smaller range covered than the azobenzene-free system, of 5 - 8.5 nm, and a similar angular range of 50 - 180°.



(a) Frame 0 (b) Frame 149

(c) Frame 399 (d) Frame 799

Figure 4.17: Stills from the 100 ns initial simulation of the cis azobenzene bar system.

Trans-azobenzene System

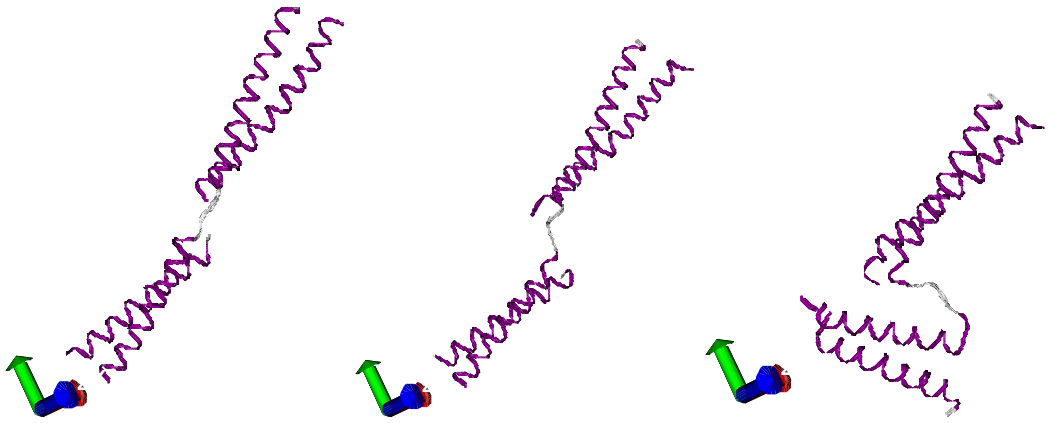
The movement of the trans system is somewhat different to the other two systems. The forced reduction in $\epsilon 2$ central structure due to the trans isomer dimensions has a significant effect. The stills of the simulation (Figure 4.18) indicate increased mobility; by 15 ns (Figure 4.18c), the central region's flexibility has already resulted in the two coiled coils lacking a shared central origin. Frames 149 and 999 have similar T-like structures; Figures 4.18c and Figure 4.18f show appearances from two viewing angles. While the linking region is not in fact a single hinge point in any of the realistic systems (unlike the concept and CG simulations), in the initial absent and cis isomer azobenzene systems, it was a (flexible) helical region with a small end-to-end separation, and predominantly behaved in ways that could be approximated by a hinge point. In the trans system, it has very little resemblance to a hinge, and is far more like two cylinders attached at one end by a length of string (Figure 4.18b).

Figure 4.15 shows that the end-to-end separations in the trans system vary from around 9 nm to 3.5 nm, and the angles created between the α-carbon atoms of three residues range from around 25 to 180°. It was at this point that it was decided, as discussed

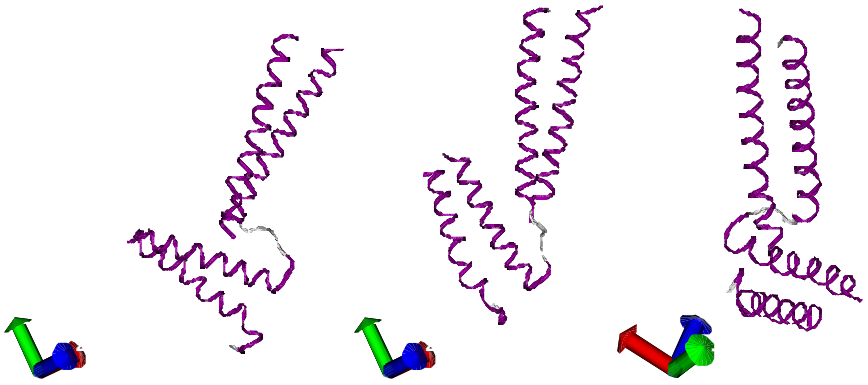
previously, that these angles were somewhat redundant as the system rarely conformed to a variable-angle V-like system. The angles calculated from the axis vectors (Section 4.5.1, Figure 4.12c) give a slightly closer representation, as they are at least correctly aligned with the coiled coils, but as the origin is not shared they still have limited use; in this method a 'T' shaped conformation and an 'L' shaped confirmation would each give a 90° angle. The angular range covered during the simulation is wide, from around 180° to 20°, a much smaller minimum angle than that of the two other systems, enabled by the flexible central region. Due to the system's angle changes being less restricted than the other systems, the end-to-end variations are less coincident with the calculated angular variations. This central region's high level of movement is consistent with the design aim of preventing helicity by using the trans conformation of the azobenzene moiety to separate the intervening residues by a distance greater than that which would allow helical secondary structure.

4.5.3 Findings from Initial Simulation

The initial study indicates that the coiled coils behave as intended. With regards to the central region, it indeed appears to be forced into a more flexible state when connected to a trans isomer of the azobenzene moiety than a cis isomer. Longer simulations are needed to ensure sampling of the full spectrum open to the system if it is given sufficient time. It is already clear that the two angles desired for the system (Section 4.1) are possible; lack of flexibility seems to not be an issue. It is unclear, however, whether the motion is sufficiently controllable; these simulations imply not. The issues encountered tend towards the opposite of what was anticipated; it was thought that the presence of the coiled coils could potentially remove flexibility from the central region, and allow the formation of a helix too rigid to be affected by the switching of an azobenzene molecule. In fact, it appears from the MD simulations that we have a system that is 'too' flexible; instead of rigid and flexible states, we have flexible and very flexible conformations. The CG simulations indicate that this may not be an insurmountable problem (discussed further in this chapter's final discussion). All that is necessary for a (non-optimally efficient) progressive system is two states with a significant difference in average end-to-end separation. Further studies into the apparent differences between the free and cis isomer systems is of interest, as it is unclear as to their origin and significance at present.



(a) Frame 0 (b) Frame 99 (c) Frame 149



(d) Frame 399 (e) Frame 599 (f) Frame 999

Figure 4.18: Stills from the 100 ns initial simulation of the trans azobenzene bar system.

4.6 MD Simulations Round 2: Increasing the Rigidity of the Bar System

The initial simulations indicated that by focussing on preventing the design of an unswitchable helix, we have instead designed a system which is able to sample a larger conformational space than is desired. As it is, the system is unlikely to fulfil the criteria of having switching bar and V-like states. Even the seemingly helical cis azobenzene system allows more flexibility than is required, and has characteristics closer to what we wished from the trans system, for a V-like conformation, rather than a bar like conformation. The originally intended V-like conformation, the trans system, is altogether too flexible. Changing the system to add more rigidity is desirable. The key question is whether a peptide system can be designed which is rigid enough for our purposes, but also switchable to a more flexible state. The cis azobenzene conformation of $\epsilon 2$ appears to be helical, but is still too flexible. Is there a sequence that is as rigid as a coiled coil but which can be switched? There are many changes that we could make to the system to try to increase the rigidity. At present, we are still in a position where the effect of single point mutations in peptides cannot be fully predicted, and hence further MD simulations are useful to guide us before synthesis of $\epsilon 2$ Version 2.0 is attempted. We have some general 'rules' which are good starting points in trying to improve the system, and we have tried to use some of these in the second round of MD simulations.

4.6.1 Excluded Ideas for Round 2

There are some ideas for rigidity increases which could be successful. However, some may not be switchable, and others are only discounted temporarily, and could be tested at a later stage.

Coiled-coil Central Region

A near certain way of increasing the rigidity of the system would be to have coiled-coil interactions continuing through the linking region of $\epsilon 2$ and the partner peptides. From the initial simulation we have done here, and significant numbers of previous studies on coiled coils, it is clear that coiled-coil regions should be sufficiently rigid. We have seen that it is the linking regions and the free ends of the peptides that are open to erratic motion. To remove the instability of the inner free peptide ends, which could still have

some freedom, even with coiled-coil interactions, a second long peptide would be needed (discussed below). We initially rejected using coiled-coil interactions in the central region of $\epsilon 2$ as we anticipated creating a helix that was too stable. We now reject the idea due to experimental findings in our group. Experiments, by ACM, using a (i, i+11) azobenzene spacing within a coiled coil, indicate that the azobenzene switch is not strong enough to disrupt a coiled coil. Further experiments using an (i, i+7) azobenzene spacing in a coiled-coil system would be desirable, but this current evidence is sufficient to reject the idea unless any contrary evidence is obtained.

A Second Long Peptide to Replace the Short Partners

This is another idea which comes to mind when considering reducing the flexibility of the system. At present, this idea is not particularly desirable for several reasons. We would need to synthesise a second ∼60 residue peptide and from experimental evidence, this is known to be difficult. $\epsilon 2$ was a complicated synthesis, while a similarly lengthed peptide attempted with standard conditions did not result in a successful synthesis. If the second long peptide follows the same design as $\epsilon 2$, including the azobenzene moiety, but without coiled-coil interactions, there is no reason to expect the linking region sequences will interact with each other, and hence the rigidity will unlikely increase. Furthermore, when switched to the trans state, although there may be a slight decrease in the conformational space open to the system due to the connecting of the inner ends of the partner peptides (for example, the conformation at 15 ns in Figure 4.18c, may be prevented due to a maximum separation of the current inner residues being imposed by the introduction of a second linking region), most of the flexibility will remain. In addition to this, there could be a future issue regarding molecular biology and the repressor feet; in order to synthesise a peptide with an azobenzene linker, one would likely wish to attach the repressor feet to other peptides, and it would be substantially more difficult to produce a ∼60 residue peptide with a repressor on each end, than to produce a single foot with a coiled-coil linker.

(i, i+11) Spacing

This idea has only been rejected due to time restrictions, as we are nearing the end of the project. We have seen that both the cis and trans systems are less rigid than we had hoped. The cis system is still able to bend even though it is the more helical state. The trans

system both disrupts the linking region structure, and has a greater spatial separation between the coiled coils than the cis system. There are several possible changes that could occur if the spacing was changed from (i, i+7) to (i, i+11). There is first the decision as to whether the central region is extended, so that there are 10 residues rather than six in the region interceding the cysteine residues, or whether the cysteines are simply shifted so they are placed more firmly in the ends of the coiled-coil regions as opposed to being in the central region. The (i, i+11) system may be slightly more restricted in its movement than the (i, i+7) system, simply due to the fact that if the trans system behaves similarly to the cis system of (i, i+7), and helicity persists as in cis $\epsilon 2$, it will likely still have the same tendency to bend more than we had intended, but the corresponding cis version may be more restricted than the trans $\epsilon 2$, due to the fact that the cis isomer has less spatially separated ends, and hence reduces the allowed separation of the coiled-coil ends. Whether this is in fact useful, or just a slightly improved version of the current system is currently undecided - it is likely that this would only be useful in combination with a central region sequence change to improve its rigidity. This may be a useful change to make small improvements to the control of the system, once the rigidity issue has been resolved.

4.6.2 Mutations of the Central Region

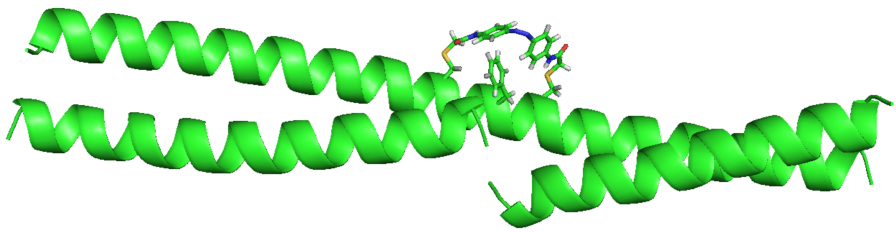


Figure 4.19: Ala27Phe Mutant of $\epsilon 2$CL, with peptides 1 and 6.

It was decided that trying the increase the rigidity of the central region using subtle changes in the system, without changing the cysteine/azobenzene locations, or introducing a coiled-coil region, was a sensible first step. As we have found that our issue is a lack of rigidity and not an unswitchable helix, we tried to make mutations towards increasing the helicity, and hopefully, consequently, the rigidity. The mutants tested were single point mutations: Ala27Phe, Gly26Ala and Asn29Ala, and a mutant with both Gly26Ala and Asn29Ala mutations. The residues were mutated using UCSF Chimera [222], and the

energy minimisation and simulations carried out again using GROMACS, in collaboration with RBS. Gly26Ala and Asn29Ala are single point mutations which will increase the number of alanines (and the length of polyalanine present) in the central region, which should increase the helicity of the region [80; 81; 82]. Ala27Phe was simulated to investigate the possibility of an interaction between a phenylalanine side chain and the azobenzene moiety stabilising the structure. The model of Ala27Phe is shown in Figure 4.19.

4.6.3 Results and Discussion

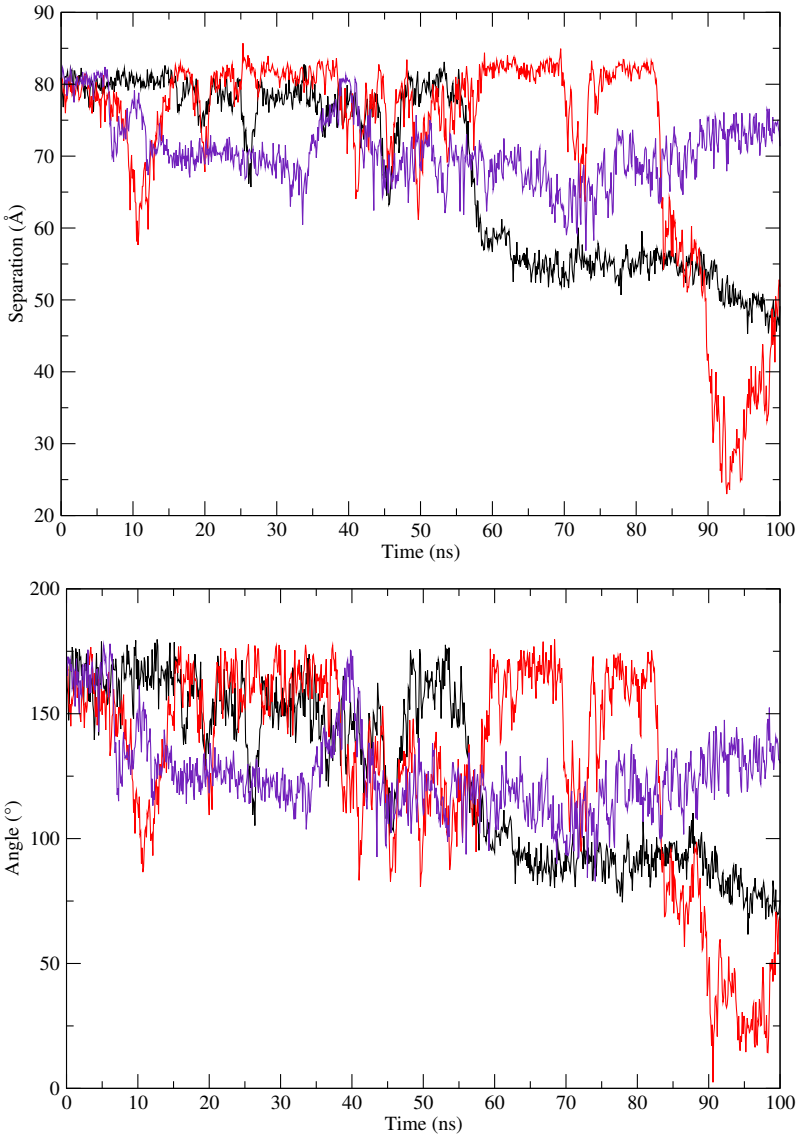


Figure 4.20: Evolution of Azobenzene-free Bar System - Runs 1-3

The second 100 ns simulation of the azobenzene-free system has a highly flexible central

Table 4.1: MD Simulation Sequences

ϵ2 variant used	**Sequence**	**Azobenzene conformation**[1]	**No. of 100ns simulations**
ϵ2	ϵ^2_{bar}CAGAANAC2$_{bar}$[3]	-	1+2[4]
ϵ2	ϵ_{bar} CAGAANAC 2$_{bar}$	cis	1+2[4]
ϵ2	ϵ_{bar} CAGAANAC 2$_{bar}$	trans	1+2[4]
Ala27Phe	ϵ_{bar} CAGFANAC 2$_{bar}$	cis	3
Gly26Ala	ϵ_{bar} CAAAANAC 2$_{bar}$	cis	3
Asn29Ala	ϵ_{bar} CAGAAAAC 2$_{bar}$	cis	3
Gly26Ala-Asn29Ala	ϵ_{bar} CAAAAAAC 2$_{bar}$	cis	3

[1] Azobenzene is attached to the long peptide via the two cysteine residues in the sequence.

[2] ϵ_{bar} is the peptide ϵ-based ϵ2 component (residues 1-23): EIAALEYEINALEQKIAALKQKN.

[3] 2$_{bar}$ is the peptide 2-based ϵ2 component (residues 32-59): NKIAALKQEIYALEQKNAALKQKIAALK.

[4] These simulations were initially carried out once for the initial study, and then repeated twice during round 2 of simulations.

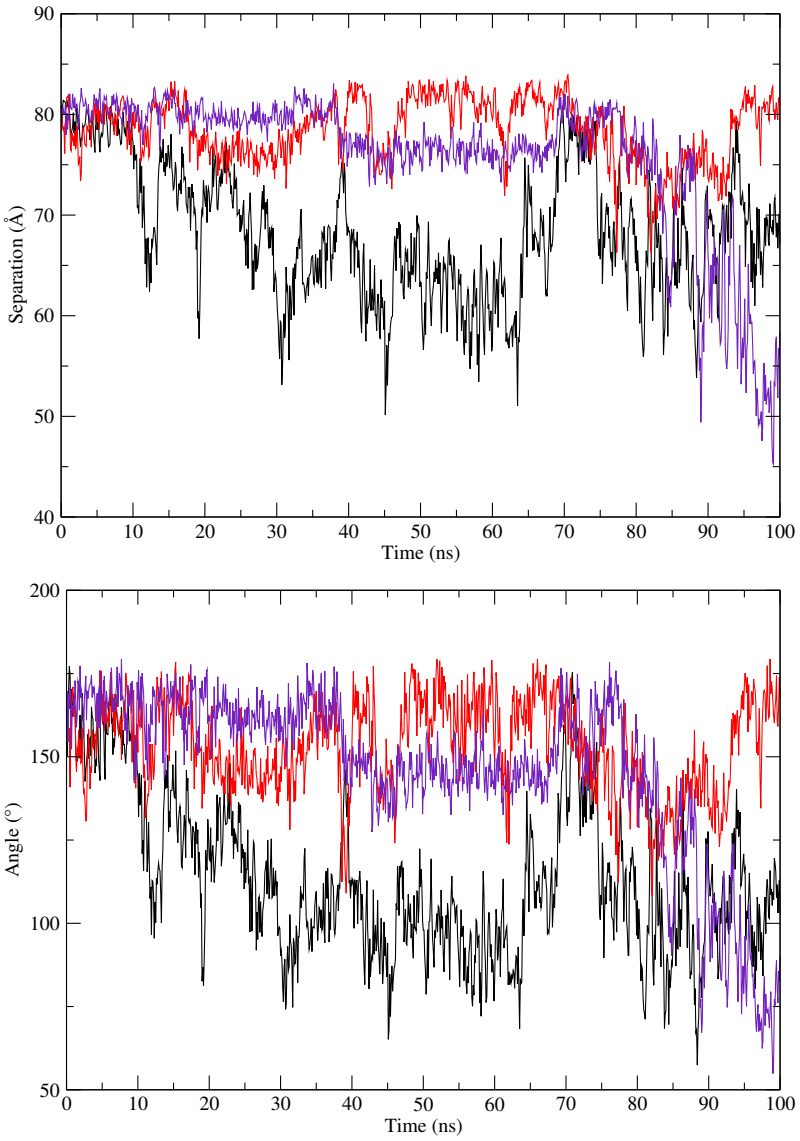


Figure 4.21: Evolution of Cis-azobenzene-containing Bar System - Runs 1-3

region, indicating that it is the peptide sequence which allows this as much as the trans azobenzene molecule. At one point, at ∼91 ns, it has near-parallel coiled coils (Figure 4.27a), dubbed a 'Π-like' conformation, having displayed much greater angular variability than any of the three initial simulations. It demonstrates the system's ability to take straight and V-like conformations, along with conformations between the previously seen V- and T-like structures (Figure 4.27b), and the new Π-like conformation. The third simulation shows no new behaviours, occupying various states between bar-like and an 80° 'V' conformation.

Nothing significant occurs in the second and third simulations of the cis azobenzene sys-

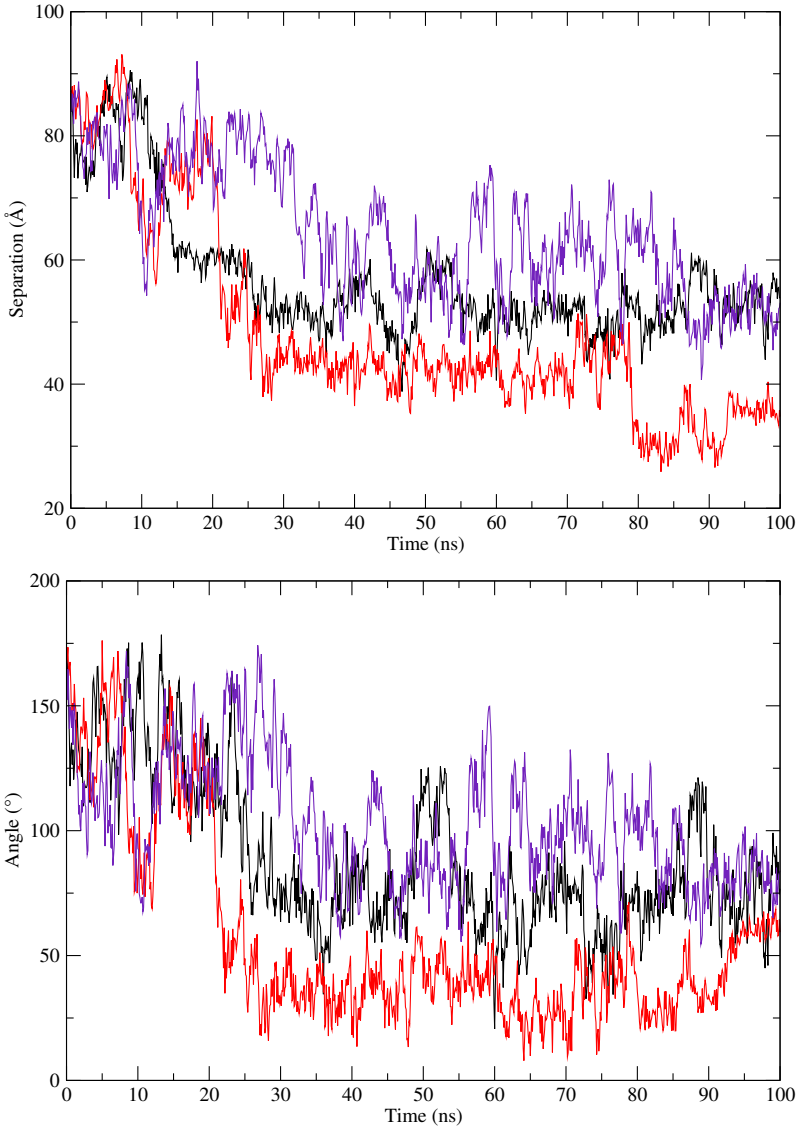


Figure 4.22: Evolution of Trans-azobenzene-containing Bar System - Runs 1-3

tem. The end-to-end separations and angles trace similar patterns, and, like the initial simulation, occupy a variety of straight and V-like states, with a range of angles from 50 to 180°.

The second simulation of the trans azobenzene system shows behaviour akin to two independent coiled coils for the first 25 ns, and afterwards a more connected system. There is a lot of uncoiling of the inner ∼7 residues of peptide 6 (Figure 4.27j) and the linking region undergoes a lot of erratic motion. The system does however, still spend a lot of time at a small angle with a slightly displaced origin. The third simulation shows a much more steady system. Representation of this simulation using a single central origin would be much more applicable than in the other two trans simulations. There is no T- or Π-like

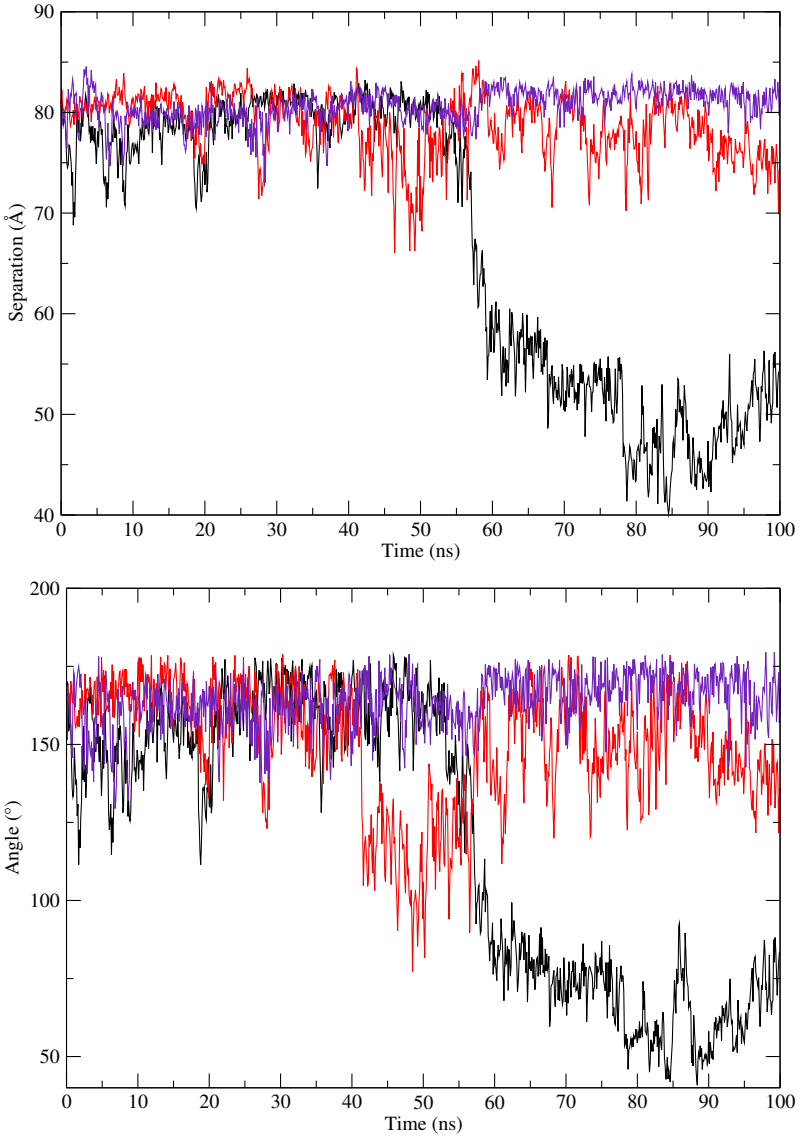


Figure 4.23: Evolution of Cis-azobenzene-containing Bar System - G26A mutant - Runs 1-3

behaviour. Closer inspection of the simulation shows that there is a crossover between the peptides in the system (Figure 4.27e, 4.27f). There is only one long peptide, and hence the peptides cannot be fully intertwined, but this occurrence may assist the stability of the system, as the fact that this does not occur in any of the other simulations hints that there may be a connection (Figure 4.27d shows a more typical arrangement of $\epsilon 2$, with no crossover).

Two of the G26A runs do very little, and spend the majority of the simulations almost completely outstretched, with angles averaging around 140 - 160°. The first run demon-

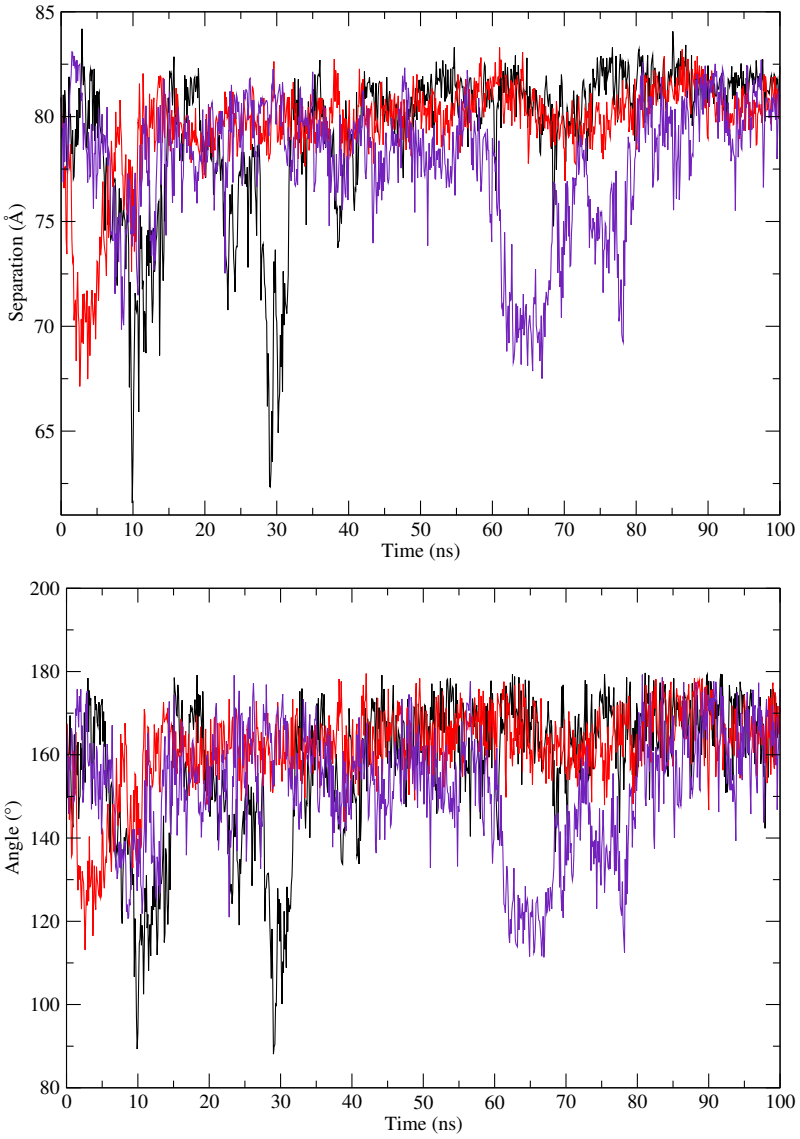


Figure 4.24: Evolution of Cis-azobenzene-containing Bar System - N29A mutant - Runs 1-3

strates, however, like its wild type counterpart, that the system is able to form V-like conformations, and shows a 40 - 180° range (∼40 - 90 Å end-to-end separation).

The N29A runs all spend the majority of their time in a straight conformation, and also display some V-like behaviour.

The G26A-N29A mutants show little unusual behaviour, again displaying oscillations between bar to V conformations. There is one occurrence of interest: in Run 1, at ∼21 ns (Figure 4.27g), the bend in the system is not due to a flexible helical region. The helix appears to continue from peptide 2 through the central region, and the distortion allowing

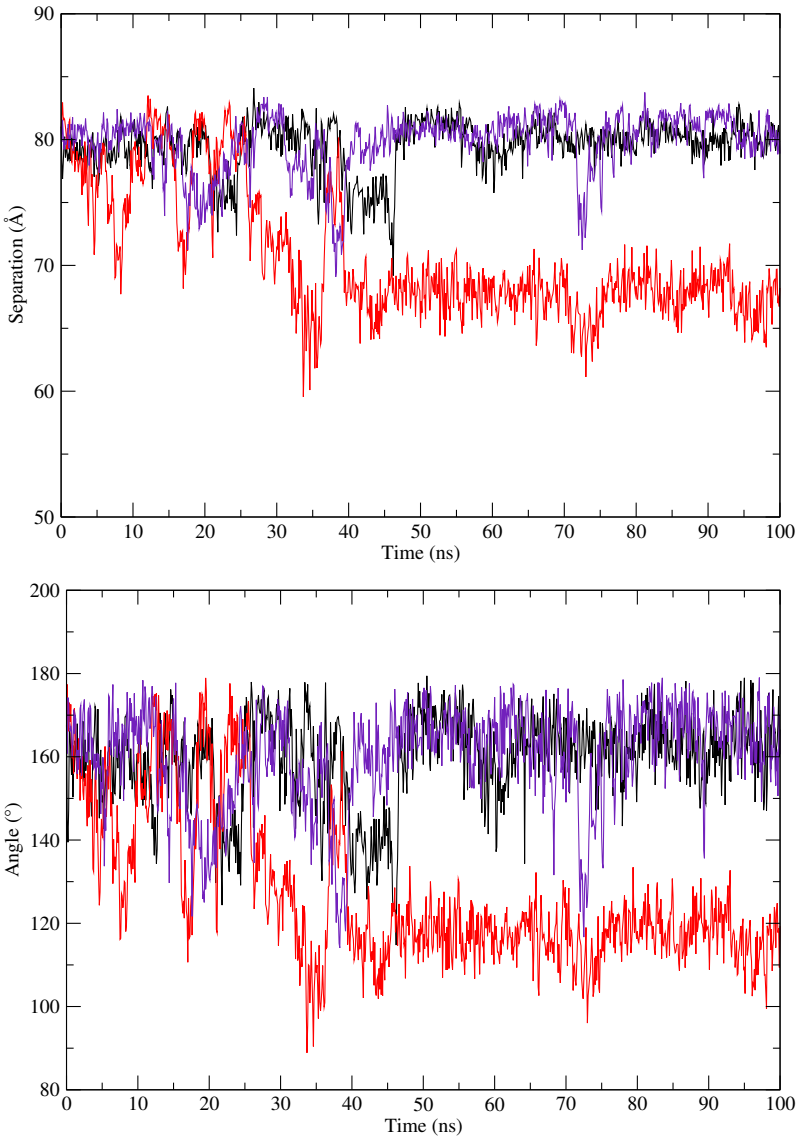


Figure 4.25: Evolution of Cis-azobenzene-containing Bar System - G26A, N29A mutant - Runs 1-3

bending is between residues 18 - 23, within the ϵ domain. This behaviour is unusual, and undesirable, but as it is not seen frequently, is unlikely to represent a statistically significant issue.

The A27F mutants, again, show very little difference in overall behaviour. The first simulation spends much of its time seemingly locked in a $\sim$100° angle V-like conformation, whilst the other two simulations display bar and V-like behaviour.

Many of the simulations show $\sim$7 residues uncoiling at the inner end of the shorter peptides (Figure 4.27c, 4.27d, 4.27i, 4.27j), in particular peptide 6.

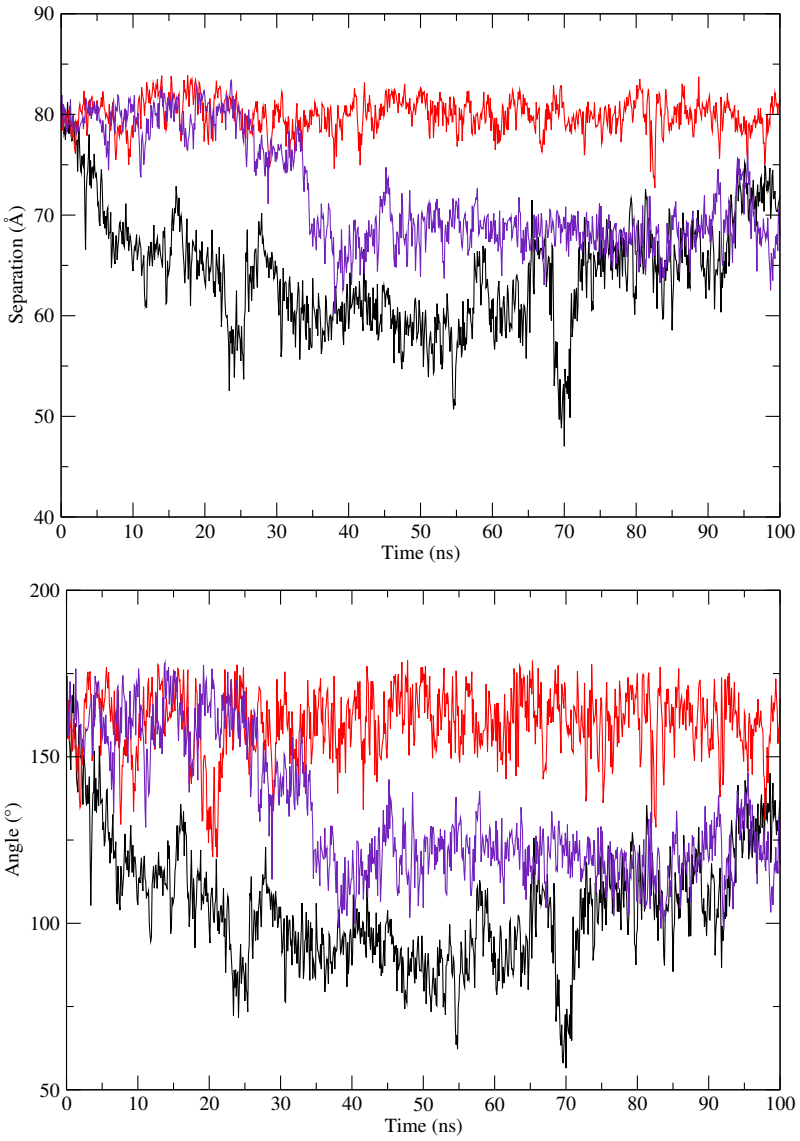
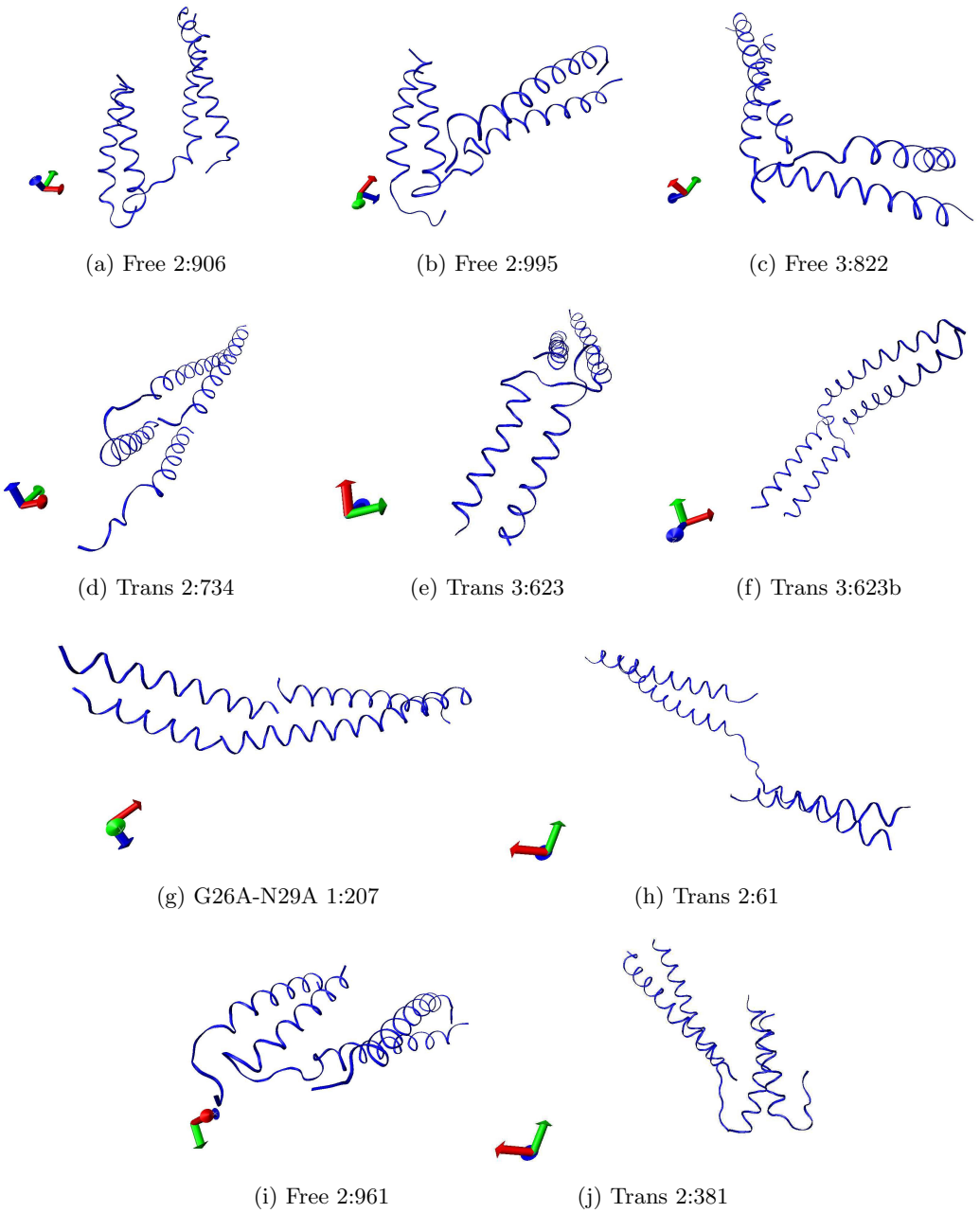


Figure 4.26: Evolution of Cis-azobenzene-containing Bar System - A27F mutant - Runs 1-3

4.7 Conclusions

With the first and second rounds of simulations, we are starting to gain a better idea of what is common behaviour for the systems. It appears that the system when unconstrained by an azobenzene molecule often displays bar and V-like behaviour, but is able to have an almost completely unstructured central region, and display accompanying erratic behaviour with no single hinge point. The trans azobenzene systems display far more erratic behaviour than the cis systems, but this behaviour is tempered somewhat in the simulation where there is some degree of 'tangling' between the peptide chains. As there



(a) Free 2:906 (b) Free 2:995 (c) Free 3:822

(d) Trans 2:734 (e) Trans 3:623 (f) Trans 3:623b

(g) G26A-N29A 1:207 (h) Trans 2:61

(i) Free 2:961 (j) Trans 2:381

Figure 4.27: Snapshots of Simulation Round 2. The labels state the peptide system ('Free' refers to the azobenzene-free system), and simulation number:frame number.

is no way we could force this behaviour, whether it is a significant effect or purely coincidental with the more hinge-like behaviour of the system is slightly redundant. Of all of the 21 simulations, it is only the trans systems, and one of the three non-azobenzene linked runs which show significantly unhelical regions. This is clearly displayed in the separations of the two cysteine residues of $\epsilon 2$ for the different runs, shown in Figure 4.28. There are several others where a few central residues are distorted, allowing bending, but

not to an extent which compromises the hinge like-behaviour of the systems. None of the mutations tested gave a sufficiently rigid performance to create a bar to V switch system. It seems a reasonable assumption however, based on the data set so far, that having a cis azobenzene molecule in the system, although allowing bending, prevents the coiled coils from separating far from each other, and so can be modelled as a hinge-like system. The mutations may add some stability, but a greater number of simulations would be required to see if there is a significant increase in rigidity in any of them compared with the wild type. As they all contain cis isomer azobenzene molecules, however, they show that in 15 100 ns simulations, the cis system does not stray from a hinge system, while both of the other systems do, even with only three 100 ns simulations each thus far. A cis azobenzene system could potentially display all of the characteristics needed for a bar motor hub. It is able to occupy a range of conformations, including those originally suggested for the two states. The purpose of the azobenzene moiety was to be able to apply a switch to the system, and although this switch does not have as great an effect as hoped in these simulations, it may still have potential if the switch is sufficient to move the system between two states which are a less-flexible state and a more-flexible (open to smaller end-to-end separations more frequently) state. Further simulations are required to give significance to the difference in average flexibility of the cis and trans azobenzene systems.

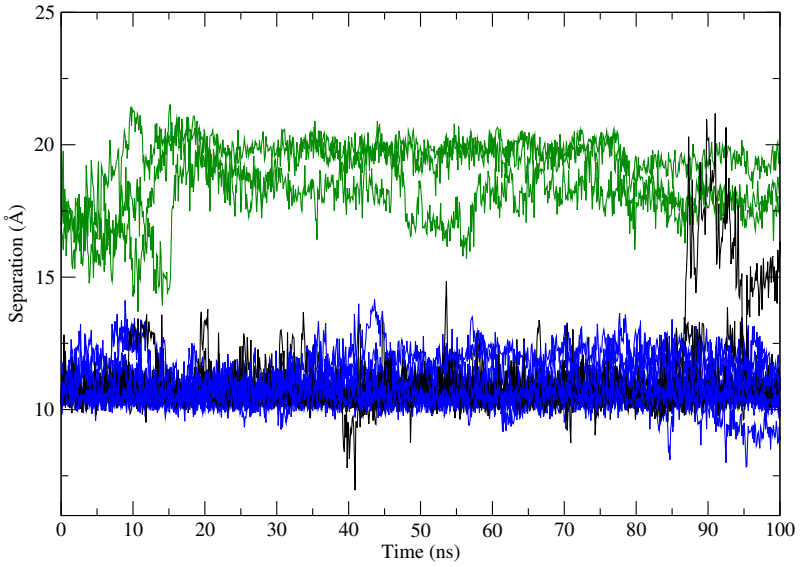


Figure 4.28: Separations between the two cysteine residues in the different MD simulation runs. The black spectra are of the non-azobenzene system, while the cis (wild type and mutants) systems are shown in blue, and the trans systems in green.

Chapter 5

Bar Motor Experiments

Alongside the MD simulations of $\epsilon 2$ and $\epsilon 2$ with an azobenzene molecule (Figure 4.7) attached (from here denoted as $\epsilon 2$CL), experiments on the system were carried out on the initial sequence of $\epsilon 2$. $\epsilon 2$ and $\epsilon 2$CL were synthesised by ACM. The experiments in this chapter were carried out by the author.

5.1 Experimental Aims

These experiments investigated the feasibility of a system such as $\epsilon 2$, 1 and 6 as a motor hub. Experiments were carried out with the following aims:

- to establish whether $\epsilon 2$ is helical, and whether it interacts with partners, by measuring the circular dichroism signals of it in isolation, and with partners.
- to measure the effect of attaching an azobenzene moiety, and the effect of this moiety when its conformation is switched, by comparing $\epsilon 2$ alone, $\epsilon 2$ plus azobenzene, and $\epsilon 2$ plus azobenzene when irradiated.
- to establish how long the azobenzene moiety remains in the second conformation after irradiation, before switching back, or whether it has to be irradiated at another wavelength to achieve this.
- to establish approximate sizes (using DLS) of the species present for samples of $\epsilon 2$ and $\epsilon 2$CL alone, with partners, and with disulphide bonded partners (peptide 1-6); for $\epsilon 2$CL, these sizes will be examined both before and after irradiation with 360nm light.

5.2 Results and Discussion

5.2.1 CD Data

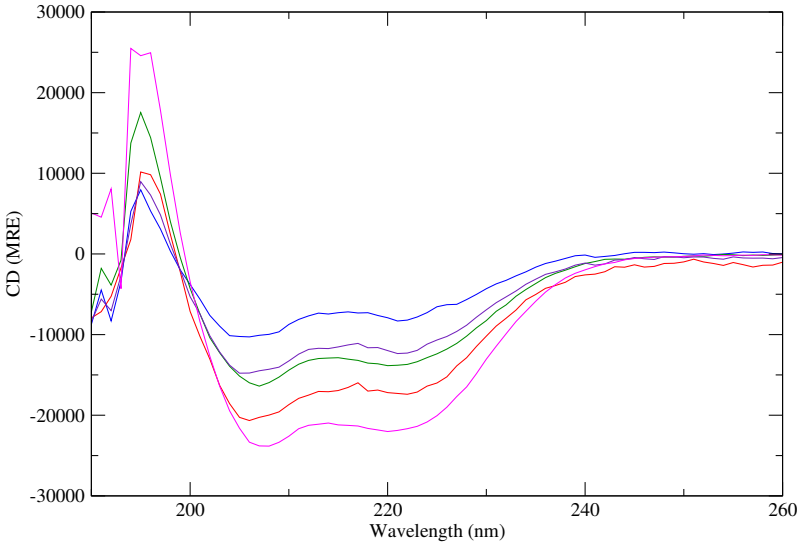


Figure 5.1: CD data of ϵ2 (magenta) compared with its nearest single analogues (though they differ by linking regions), ϵ (red) and 2 (blue), and with those single analogues in solution together (green). The theoretical spectrum of ϵ,2 is also shown (purple). Measurements were taken at 20°C at 20μM concentration in PBS buffer (pH 7.4).

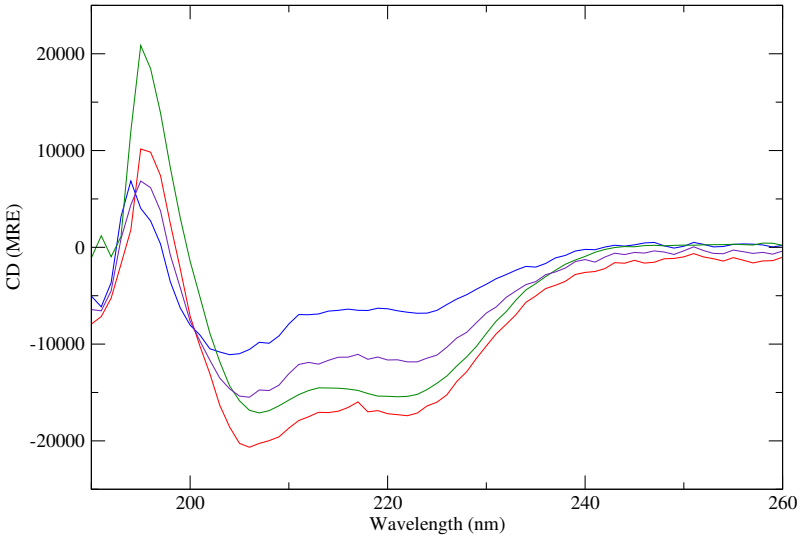


Figure 5.2: CD data of ϵ,6 (green). The spectra of the individual peptides are also shown, ϵ (red), peptide 6 (blue), along with the theoretical spectrum of the mixture (purple). The other coiled-coil pair in the bar motor, (1,2), is shown in Figure 3.4. Measurements were taken at 20°C at 20μM concentration in PBS buffer (pH 7.4).

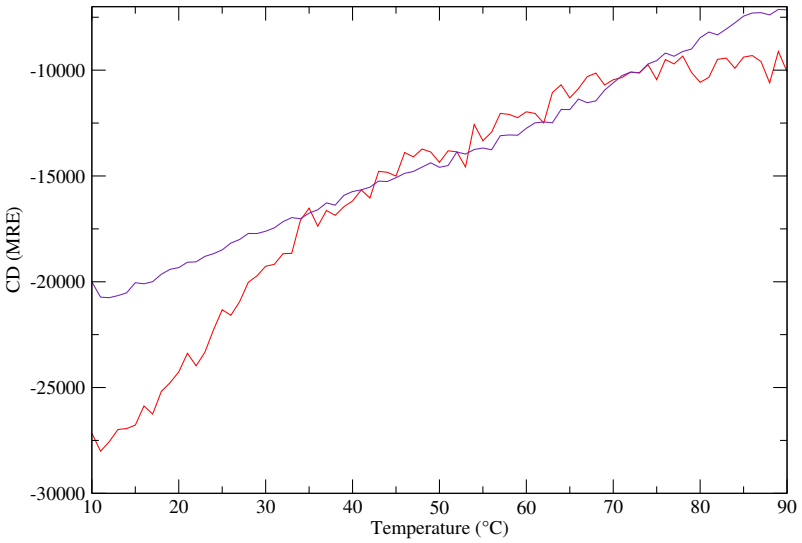


Figure 5.3: CD melt spectra of ϵ (red) and ϵ, 6 (purple). Measurements were taken at 10-90°C at 20μM concentration in PBS buffer (pH 7.4).

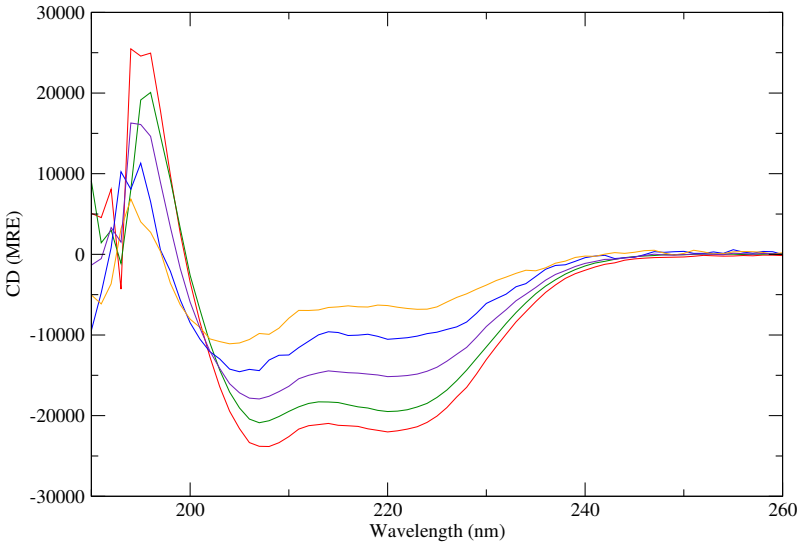


Figure 5.4: CD spectra of single peptides ϵ2 (red), 1 (blue), 6 (orange), compared to a mixture of ϵ2, 1 and 6 (green). The theoretical spectrum of ϵ2, 1, 6 is also shown (purple). Measurements were taken at 20°C at 20μM concentration in PBS buffer (pH 7.4).

ϵ2 appears to be more helical than its individual contributing peptides and the mixture ϵ,2. ϵ,2 itself appears to be slightly more helical than predicted for a system where peptides ϵ and 2 were not interacting (this theoretical spectrum is shown in Figure 5.1). This is not surprising given the interactions seen between the non-designed pairs of Chapter 3. The ϵ,6 mixture also appears more helical than predicted by the non-interacting theoretical spectrum (Figure 5.2). The melting temperature of ϵ,6 is 69.5 $\pm$ 1.1°C (spectrum

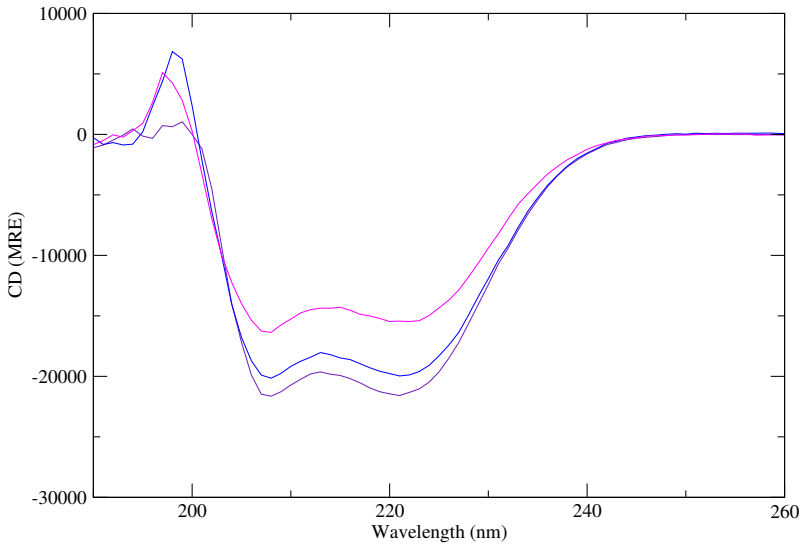


Figure 5.5: CD spectrum of ϵ2CL (blue), compared to ϵ2CL mixed with 1 and 6 (purple), and ϵ2CL with disulphide-bonded 1-6 (magenta). Measurements were taken at 20°C at 100, 80 and 50μM concentration respectively, in PBS buffer (pH 7.4).

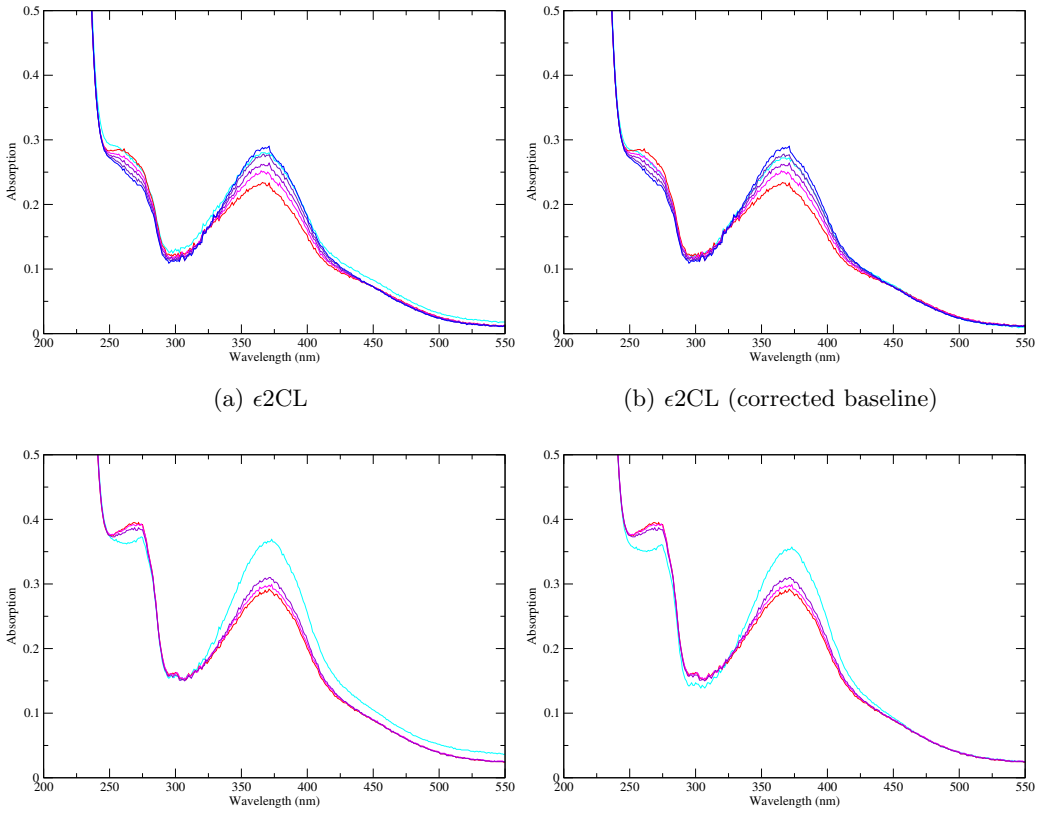
shown in Figure 5.3).

Figure 5.4 shows the CD spectrum of the ϵ2, 1, 6 mixture along with those of its contributing peptides and the theoretical spectrum calculated from them. The mixture of ϵ2, 1, 6 is again more helical than predicted by the theoretical spectrum, indicating that hetero-oligomeric interactions are present. The mixture, and ϵ2 alone display clear characteristics of helical structures, with minima around 208 and 222 nm.

Due to the change in the UV spectrum of ϵ2 when attached to an azobenzene moiety, the usual means of concentration determination was not possible. Ideally, the extinction coefficient of the new peptide ϵ2CL would be determined, but the amount of product after reactions and multiple purification stages made this impossible. As there should not be a major change in the CD signal of the peptide upon reaction with the azobenzene moiety, a comparison of the ϵ2 and ϵ2CL CD spectra was used to estimate the concentration of ϵ2CL. For this reason, there is no comparison between the CD spectra of ϵ2 and ϵ2CL, as we have had to assume for the present time that they are approximately the same. The shape of the spectra did indeed indicate that this was not a blatantly inaccurate assumption. As shown in Figure 5.5, all three samples containing ϵ2CL appear to have CD spectra that are characteristic of helical species. ϵ2CL and ϵ2CL,1,6 appear to be significantly more helical than ϵ2CL,1-6. Upon irradiation with 360 nm light for 30 minutes, CD, UV and DLS measurements were taken to compare with those taken before irradiation. The

CD spectra did not display a visible change, and only showed minor changes comparable to those between repeats of a single sample. This is unsurprising, as under 10 of the 59 residues of the peptide would be expected to be affected if switching occurred, and full conversion is also not expected [196].

5.2.2 UV Data



(a) ϵ2CL

(b) ϵ2CL (corrected baseline)

(c) ϵ2CL, 1, 6

(d) ϵ2CL, 1, 6 (corrected baseline)

Figure 5.6: UV spectra of crosslinked ϵ2 (37μM, in PBS buffer, pH 7.4) before irradiation (cyan), and recovery after irradiation at 360 nm for 15 minutes. The red, magenta, violet, indigo and blue spectra are taken at t, t+28, t+50, t+77 and t+102 minutes, where t is 5 minutes after irradiation.

The UV spectrum of ϵ2CL and ϵ2CL with both 1,6 and 1-6 are shown in Figures 5.6 and 5.7. Figure 5.6 shows preliminary data of the system prior to and after 15 minutes of irradiation with 360 nm light. Figure 5.7 shows the three ϵ2CL-containing samples, after irradiation for 30 minutes at 360 nm. In the preliminary data sets, both ϵ2CL and ϵ2CL,1,6 show a significance decrease in the absorption peak in the region of 360-370 nm. The ϵ2CL sample, perplexingly, appears to slightly overshoot the pre-irradiation spectrum

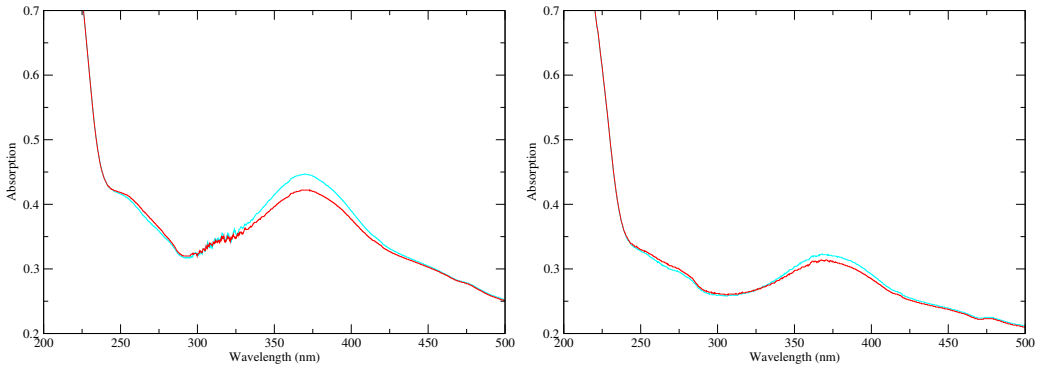
after around 100 minutes of recovery. The recovery of ϵ2CL,1,6 was only tracked for 25 minutes, but also indicates a recovery time in the region of 100 minutes. For the later data of the systems, the switch appears to be less pronounced for the ϵ2CL and ϵ2CL,1-6 samples, even though the irradiation time was increased to 30 minutes. The ϵ2CL spectrum appears to be behave in a similar manner to previously seen, while the ϵ2CL,1,6 sample shows a shift in its absorption peak. The ϵ2CL,1-6 sample, as it at half the concentration of the ϵ2CL sample, appears to have undergone a similar change to the ϵ2CL species. From the data shown in Figure 5.7, one would assume that a more significant change has occurred in the ϵ2CL,1,6 sample. This would be a good outcome as this is one of the potential Bar motor systems. The fact that the spectrum returns to almost the same spectrum as its pre-irradiation state after an hour of recovery implies that the peak shift observed is not a measurement error. ϵ2CL,1,6 may show a more pronounced switching than the other systems due to its having greater freedom in terms of being switched; the ϵ2CL will be resistant to switching if it is forming a hairpin like structure, while ϵ2CL, 1-6 would be expected to be less flexible than its non-disulphide bonded counterpart. The fact that this is not seen in the preliminary data, and that the percentage of switching seems to be highly variable indicates that optimisation of the switching protocol is required, but there are clear signs of the azobenzene moiety inducing changes to the system.

5.2.3 DLS Data

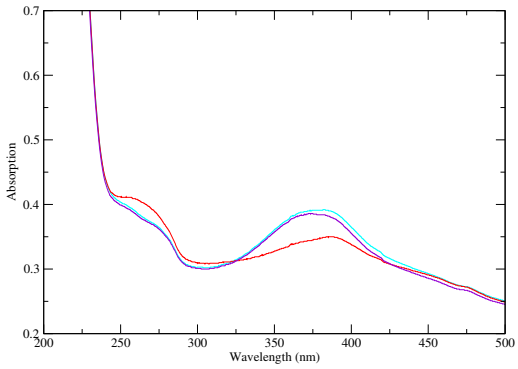
DLS data was taken for the long peptide alone (ϵ2), mixed with the two partner peptides (1,6), and mixed with the disulphide-bonded pair of partner peptides (1-6), at 100μM per peptide. The same measurement and fitting methods were used as in Chapter 3. These three experiments were then repeated using the long peptide ϵ2 linked via its cysteine residues to the azobenzene moiety (Figure 4.7), known as ϵ2CL. Due to limited product, these were carried out at 100μM for ϵ2CL alone, at 80μM per peptide for ϵ2CL with peptides 1 and 6, and 50μM per peptide for ϵ2CL mixed with 1-6.

ϵ2

ϵ2 was a difficult data set. There were 10 sets out of the 50 taken which had double peaks, and many of the peaks were less Gaussian shaped than seen in other samples, resulting in more sets with a standard deviation greater than 0.08. It seemed sensible to keep this cut off, however, as the shape of the peaks implied the possibility of a double peak



(a) ϵ2CL (100μM) (b) ϵ2CL, 1-6 (50μM)



(c) ϵ2CL, 1, 6 (80μM)

Figure 5.7: UV spectra of the three ϵ2CL-containing samples (in PBS buffer, pH 7.4) before (cyan) and after 30 minutes of irradiation at 360 nm (red). The ϵ2CL, 1, 6 spectrum also shows the sample an hour after irradiation (the purple spectrum was taken 56 minutes after the red spectrum).

mostly concealed by the strength of the more dominant peak. Excluding the results with a standard deviation greater than 0.08, and spectra with two peaks in the region of interest (ROI) left 23 data sets, and gave a size of 3.92 $\pm$ 0.21 nm. Testing with other cut offs gave results which were the same within the error found (Table 5.1).

ϵ2, 1, 6

For this data set, only one spectrum was excluded using a cut off of 0.08 standard deviation. Six spectra were excluded when double peaked spectra were removed (only one of which was excluded by the 0.08 cut-off). With a cut off of 0.08 and no double peaks, an average of 4.83 $\pm$ 0.05 nm was found, from 44 measurements.

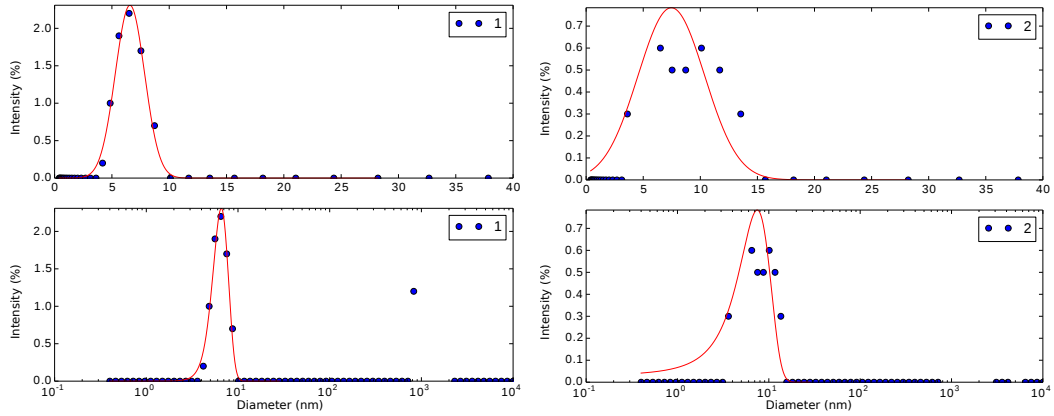


Figure 5.8: An example (left) of a good (included in the averaging process) and (right) bad fit (discarded) to two DLS data sets (runs 1 and 2) for the ϵ2, 1, 6 sample.

ϵ2, 1-6

All 50 data sets had one peak between 1 and 10 nm. No data sets were excluded for having a double peak, nor a greater than 0.08 standard deviation. An average size of 4.48 $\pm$ 0.06 was found from all 50 data sets. An average size of 4.44 $\pm$ 0.06 was found if one data set was excluded, which, with a σ of 0.053, was significantly greater than all of the other standard deviations, but using the same fitting procedure of all of the other sets, there was no need to exclude any data.

ϵ2CL

This sample was carried out at 100μM concentration, and yet of the initial set of measurements taken, a significant proportion (over 30%) showed no useful information. Extra measurements were performed, and of 144 runs, 61 showed peaks in the region of interest. Of these, 8 were excluded due to the presence of double peaks. Of the remaining 53 sets, a hydrodynamic diameter of 7.06 $\pm$ 0.11 was found. After irradiation for 30 minutes at 360 nm, 50 measurements were taken and all 50 contained information below 30 nm. The removal of double peaked sets was more difficult for this sample, as, while five sets had clearly excludable peaks, 9 had two slightly separated peaks in the region. The fact that these 9 generally had one peak located above and below the single peak of the other spectra in the set implied that they were likely products of the software's fitting process, and were discarded. From 36 data sets, a hydrodynamic diameter of 7.44 $\pm$ 0.24 nm was found.

ϵ2CL, 1, 6

Again, the initial 50 measurements taken contained few useful measurements of the species present, and hence additional sets were taken. Out of 182 measurements, only 59 had peaks in the region of interest. Of these 59, 17 were excluded for failing the selection process (one of the sets included and one of the sets excluded are shown in Figure 5.8). An average hydrodynamic diameter of 6.70 ± 0.10 nm was found from 42 measurements. After irradiation, 50 measurements were taken. 18 of the data sets had no peaks below 700 nm. The other 32 sets had one peak which fit well to a Gaussian curve. An average hydrodynamic diameter of 6.43 ± 0.18 nm was found from 32 data sets.

ϵ2CL, 1-6

Due to sample constraints, this sample was run at a lower concentration, of 50μM per peptide. The initial 50 measurements of ϵ2CL, 1-6 prior to irradiation had many data sets with no peaks of a non-contaminant origin. 150 measurements were carried out in total, and of those, 80 had peaks in the region of interest. 15 of these sets were then removed by the selection criteria of no double peaks, and standard deviations being below 0.08. For the remaining 65 sets, a hydrodynamic diameter of 6.05 ± 0.09 nm was found. After irradiation, 50 measurements of the sample were taken. Of these, only one data set had no peaks in the region of interest. 11 sets were removed for failing to fulfill the selection criteria. From 38 data sets, an average hydrodynamic diameter of 6.08 ± 0.10 nm was found.

Discussion

As discussed in Sections 2.3.3 and 3.3.2, we would expect a coiled-coil dimer made up of two ∼30-residue peptides to have a hydrodynamic diameter ∼4 nm. We would therefore also expect the bar system (ϵ2,1,6 and variants) to have a similar size if it was sufficiently flexible, and energetically favoured a structure such as that shown in Figure 4.27a. If the system is fixed in a more extended structure, such as those shown in Figure 4.10, we would expect a hydrodynamic diameter approximately twice that size, of ∼8 nm. Here, it appears that the three samples containing ϵ2 are close to a 4nm diameter (Table 5.1), while the three samples containing ϵ2CL (particularly ϵ2CL and ϵ2CL,1,6) have diameters which are larger, nearer to 8nm (Table 5.2).

Table 5.1: DLS Averages: Bar Motor $\epsilon 2$ Samples

Sample	**Criteria**	**Sets used**	**Average (nm)**	**Standard Deviation (nm)**	**Standard Error (nm)**
$\epsilon 2$	All data sets	50	4.14	1.01	0.14
	Double peak sets removed	40	4.05	0.91	0.14
	$\sigma < 0.08$	27	3.92	0.99	0.19
	$\sigma < 0.1$	41	4.11	1.02	0.16
	No double peaks, $\sigma < 0.08$	23	3.92	1.02	0.21
	No double peaks, $\sigma < 0.1$	34	4.06	0.96	0.17
$\epsilon 2$, 1, 6	All data sets	50	4.86	0.33	0.05
	Double peak sets removed	44	4.83	0.32	0.05
	$\sigma < 0.08$	49	4.84	0.31	0.05
	No double peaks, $\sigma < 0.08$	44	4.83	0.32	0.05
$\epsilon 2$, 1-6	All data sets	50	4.48	0.43	0.06
	Double peak sets removed	50	4.48	0.43	0.06
	$\sigma < 0.08$	50	4.48	0.43	0.06
	$\sigma < 0.05$	49	4.44	0.40	0.06
	No double peaks, $\sigma < 0.08$	50	4.48	0.43	0.06

Table 5.2: DLS Averages: Bar Motor ϵ2CL Samples

Sample	**Criteria**	**Sets used**	**Average (nm)**	**Standard Deviation (nm)**	**Standard Error (nm)**
ϵ2CL	All data sets	61[1,2]	6.93	0.92	0.12
	No double peaks, $\sigma < 0.08$	53	7.06	0.81	0.11
ϵ2CL post-irradiation	All data sets	50	7.44	1.75	0.25
	No double peaks, $\sigma < 0.08$	36	7.44	1.43	0.24
ϵ2CL, 1, 6	All data sets	59[1,3]	6.72	0.89	0.12
	No double peaks, $\sigma < 0.08$	42	6.70	0.62	0.10
ϵ2CL, 1, 6 post-irradiation	All data sets	32[4]	6.43	1.00	0.18
ϵ2CL, 1-6	All data sets	80[1,5]	6.14	0.83	0.09
	No double peaks, $\sigma < 0.08$	65	6.05	0.69	0.09
ϵ2CL, 1-6 post-irradiation	All sets	49[6]	6.07	0.75	0.11
	No double peaks, $\sigma < 0.08$	38	6.08	0.64	0.10

[1] Initial 50 measurements had low count of peaks in ROI, so further sets were taken.

[2] 61 out of 144 sets had peaks in the ROI.

[3] 59 out of 182 sets had peaks in the ROI.

[4] Only 32 out of 50 data sets had a peak in the ROI; no data was excluded due to double peaks or bad fits.

[5] 80 out of 150 sets had peaks in the ROI.

[6] One of 50 data sets had no peak in the ROI.

Table 5.3: DLS Averages (time spaced): Irradiated ϵ2CL

Criteria	**Sets used**	**Number of sets**	**Average (nm)**	**Standard deviation (nm)**	**Standard Error (nm)**
All data sets	1-10	10	7.20	2.31	0.73
	11-20	10	7.99	1.88	0.60
	21-30	10	7.23	1.34	0.42
	31-40	10	7.73	1.45	0.46
	41-50	10	7.06	1.81	0.57
No double peaks, $\sigma < 0.08$	1-10	6	7.15	1.18	0.48
	11-20	7	7.45	1.90	0.72
	21-30	9	7.50	1.07	0.36
	31-40	8	7.90	1.33	0.47
	41-50	6	7.01	1.92	0.78

Table 5.4: DLS Averages (time spaced): Irradiated ϵ2CL, 1 ,6

Criteria	**Sets used**	**Number of sets**	**Average (nm)**	**Standard deviation (nm)**	**Standard Error (nm)**
All data sets[1]	1-10	7	6.50	0.68	0.26
	11-20	5	6.79	1.14	0.51
	21-30	8	6.62	0.61	0.22
	31-40	6	6.02	1.34	0.55
	41-50	6	6.19	1.35	0.55

[1] 32 data sets, not 50, were used in total, due to some sets having no peaks. Therefore, for example, the 1-10 set measurement is made up of the average of 7 sets, as three were blank.

Table 5.5: DLS Averages (time spaced): Irradiated ϵ2CL, 1-6

Criteria	**Sets used**	**Number of sets**	**Average (nm)**	**Standard deviation (nm)**	**Standard Error (nm)**
All data sets	1-10	10	5.79	1.06	0.33
	11-20	10	6.12	0.64	0.20
	21-30	10	6.32	0.47	0.15
	31-40	10	5.84	0.82	0.26
	41-50	9[1]	6.29	0.58	0.19
No double peaks, $\sigma <$ 0.08	1-10	9	5.56	0.80	0.27
	11-20	8	6.23	0.39	0.14
	21-30	8	6.48	0.35	0.12
	31-40	5	5.70	0.38	0.17
	41-50	8	6.35	0.59	0.21

[1] One data set had no peaks in the ROI.

5.2.4 Discussion: ϵ2 With and Without Partners

The DLS data of ϵ2 showed a species of approximately 4 nm. As ϵ2 is a 59-residue peptide, with seven coiled-coil heptads and ten other residues (the linking region), we would expect a length of 8-9 nm if the peptide was forming a straight helix. As we are measuring the hydrodynamic diameter and not the true length, the values seen may vary somewhat from this. However, a size half of this, as seen for ϵ2 in isolation, implies that the peptide has formed compact structures; it may have formed a hairpin, with the peptide 2-based heptads interacting with the ϵ-based heptads. Alternatively, the peptide could be forming some intermediate between the two extremes of hairpin and bar, but as this is a single peptide with the hydrophobic *a* and *d* residues typical of a coiled-coil peptide, the preferred state is likely to bury these residues; forming a hairpin is the most obvious choice given the size measured. The flexibility seen in the MD simulations of the azobenzene free system of ϵ2, 1 and 6 supports this hypothesis; the linking region does seem to be sufficiently unstructured in the absence of an azobenzene moiety to allow a hairpin (or near-hairpin)

to form (the closest resemblance to a hairpin seen in the simulations is shown in Figure 4.27). The simulation was carried out for the peptide with partnering peptides without a linking disulphide bond and the V-like, T-like and Π-like conformations seen would all give similar sizes if observed by DLS. The ϵ2, 1, 6 and ϵ2, 1-6 DLS samples both give similar results to the isolated peptide, with hydrodynamic diameters of 4.83 ± 0.05 nm and 4.44 ± 0.06 nm respectively. For these systems, the peptides may be less driven to form a hairpin, though from the DLS data alone, it is difficult to speculate on the structures forming. The values again rule out a straight coiled-coil-like structure, and do not rule out the possibility of hairpin formation. The simulations indicate that these systems (as the disulphide bonded system is unlikely to have greatly different behaviour to the ϵ,1,6 system) may be forming intermediates such as the V-, T- and Π-like structures. The fact that both systems give slightly larger diameters than the single peptide supports this. We had expected that the disulphide bond between peptides 1 and 6 could possibly cause the two helical regions to be brought closer together, but this technique is not sensitive enough to measure such a change, especially given that our system is not spherical, and hence the hydrodynamic diameter does not represent the true length, only an approximation. Even if this change were measurable, the fact that intermediates between a hairpin and a straight helical structure are likely the present structures, this change would not be pronounced enough to be observed.

5.2.5 Discussion: ϵ2 vs. ϵ2CL

Unlike the change between a lone ϵ2 and ϵ2 with the addition of free and disulphide bonded partners, the addition of the azobenzene moiety to the system does appear to have a significant effect on the species size. It appears that the ϵ2CL peptide is forming a larger structure than the ϵ2 peptide. This indicates that at least some of the ϵ2CL peptides may be making a homodimer (or higher order structure) to bury their hydrophobic residues, as opposed to forming hairpins, as this is surely preferable than exposing a number of hydrophobic residues. This measurement does not rule out the presence of some hairpins, but indicates that there has been a shift in the population towards a larger structure.

5.2.6 Discussion: ϵ2CL With and Without Partners

Like ϵ2, ϵ2CL does not seem to undergo a major size change upon addition of partners. ϵ2CL has a hydrodynamic diameter of 7.06 ± 0.11 nm, while ϵ2CL, 1, 6 has a radius of

6.70 ± 0.10 nm, and ϵ2CL, 1-6 a diameter of 6.05 ± 0.09 nm. The simulations indicated that the cis isomer systems tended to spend more time in a near-bar state than the trans systems. However, both were able to occupy both near-bar and more bent V-like states. These DLS results indicate that the structure of the peptide with an azobenzene attached is more rigid without irradiation than expected, or predicted by the MD simulations. The fact that there is a significant change between the ϵ2 and ϵ2CL peptides implies that the azobenzene is successfully connected to ϵ2CL at two points, and is able to inflict change on the system. There is no obvious reason for the size of ϵ2CL, 1, 6 to be greater than that of ϵ2CL, as both are clearly in extended states. The fact that ϵ2, 1-6 results in a hydrodynamic diameter 1 nm smaller than ϵ2 may show the effects of the disulphide bond restricting the flexibility of the central hinge, but all of these changes seen between the ϵ2CL samples are small compared with the change from an azobenzene-free system.

5.2.7 Discussion: Irradiation of ϵ2CL

Irradiation of the samples does not result in a large effect as measured by DLS. From UV data, it is clear that the system is undergoing some level of azobenzene conformational switching (Figure 5.7). It appears that the switching is reversed without stimulation within ∼100 minutes. As it takes ∼43 minutes to take 50 DLS measurements, we would expect half of the switched molecules to have reverted back to their original state within this time frame. In order to ensure that no trend was missed, from the DLS data sets of the three ϵ2CL-containing samples, the average size found for each ten data sets was found, so that the average hydrodynamic diameter for the samples within ∼15 minutes of irradiation was measured, and in 8 minute steps after this (Tables 5.3 - 5.5). This data showed no progressive increase or decrease in size over time, and the deviations found did not indicate an increase or decrease in variability over time. It is unclear whether there is significance in the fact that the partner-free system (ϵ2CL) showed an increase in size upon irradiation while the other two (ϵ2CL, 1-6 and ϵ2CL, 1, 6) showed almost no change and a slight decrease, respectively. The expectation was to see a more compact structure prior to irradiation, and potentially a larger (more rigid) structure upon irradiation. As there is already a population of larger structures prior to irradiation, it is more difficult to observe the effect of the irradiation.

5.3 Conclusions

In this chapter, we have established that the introduction of an azobenzene molecule has a significant effect on the size of the $\epsilon 2$ peptide. The CD data of the system indicates that interaction between the peptides is present, and the UV data of the azobenzene-containing systems shows that the system's irradiation does induce a change. However, the effect of the introduction of the azobenzene moiety on the system was greater than predicted by the MD simulations. Determination of the extinction coefficient, allowing the collection of further information on the differences between the structures of $\epsilon 2$ and $\epsilon 2$CL, is desirable. Further work into the optimisation of azobenzene switching would be beneficial to ensure optimal efficiency of any future systems. This would give a clearer view of the effect of irradiation on the system; at the moment we are observing a larger structure that the compact $\epsilon 2$ structures prior to irradiation, and so changes upon irradiation are less clear. The use of techniques which are sensitive to individual molecules are needed to gain a better insight into the system's behaviour. FRET, using fluorescently tagged peptides, with fluorescent tags placed at the extremities of the system (for example, at the two ends of peptide $\epsilon 2$), would be a useful technique, as we would be able to determine the separation of the fluorescent tags, and hence the ends of the peptide system. The fact that such a significant change occurs between the azobenzene-free and azobenzene-bound peptides presents a possible future direction; the switchable attachment and detachment of the azobenzene moiety (with one end permanently attached and the other detachable) could be worth investigation. If possible, it appears that the two resulting structures would have significant size differences as desired in the Bar motor design.

Chapter 6

Conclusions

The experimental study of the components and formation of two hub systems has been carried out, along with simulations of the second system.

For the Tumbleweed hub, there is strong evidence that the design specification of producing three orthogonal coiled-coil dimers has been successful. The data gained on the system indicates that the pairings are sufficiently specific that a hub with three different coiled-coil pairs, and hence three different feet in the final motor hub, should form. The linkages provided at present by disulphide bonding could be replaced in order to form more robust links between the three hub legs, and this would hopefully widen the chemical environment in which the hub can successfully function, including those required for the preferential DNA-binding of the different foot domains. There are many possibilities for this change, including the synthesis of three long peptides which each span the distance between two feet of the hub, or by native chemical ligation of the peptides.

For the Bar motor, coarse-grained simulations indicate that a motor can still be progressive even if the switch in the system is not between two states of fixed angles, but one fixed state and one flexible state. The MD simulations of the system indicated that both the cis and trans azobenzene states allow some level of flexibility in the peptides, but that the trans state spends more time in non-bar-like states. If there is no difference between the foot separations open to the two states, we would still likely have a motor which could move along a track but directionality would be lost and the processivity would be lowered. The simulations indicated, however, that there is a difference in the average conformation of the two systems, and hence some level of directionality could be introduced to the system using the two azobenzene isomers and two varying binding site separations, by choosing one binding site separation which is more regularly accessed by each of the

isomer states of the azobenzene-linked peptides.

The experimental studies of the system, however, indicated that the ϵ2CL samples all contained structures which were much closer to bar-like states than more compact conformations; the presence of the azobenzene moiety appears to have a greater effect on the structure than expected from MD simulations. It would therefore be desirable to investigate further the source of the current discrepancies between the experimental observations and the MD results. The choice of force field and water model for the simulations have been used to predict several experimental peptide systems successfully [215; 216; 217] and hence it would be interesting to establish whether it is an issue with these, or perhaps the representation of the azobenzene molecule in the models, which is the source of the mismatch. Further experimental work and corresponding MD simulations could assist in this. Irradiation did not indicate a switch from bar-like to a significantly more compact structure for the levels of irradiation achieved. For the current system, optimisation of the azobenzene molecule switching, by increasing the efficiency of transfer through the irradiation method, is important to ensure that a larger proportion of the molecules are effected by the irradiation, and can be studied further to determine the difference between the conformations of peptides attached to the two different isomers of the azobenzene moiety. Following this, studying the effects of different linking regions on the levels of switching achieved is another possibility. It may be possible to simply use the ϵ2CL peptide as the hub, as it appeared to undergo a size change of 0.5 nm after irradiation (which was not covered by the standard error of the two measurements), which may be sufficient to exploit. The ϵ2 peptide, without azobenzene, with, and without peptide partners formed significantly more compact structures that the azobenzene-containing systems. This clear result, that even the unpaired ϵ2CL system is forming larger structures than the azobenzene-free systems indicates that the introduction of the azobenzene is successfully counteracting the tendency of the peptides to bury their hydrophobic residues with the most convenient partner (in this case, the peptide folding up to interact with the other half of itself). The system must be forming dimers (or higher order oligomers) instead of hairpins to bury these residues in a way allowed by the restrictions enforced by the azobenzene moiety. Based on the current results, exploiting the difference between the azobenzene-free and azobenzene-containing systems may be a good future direction to follow, with the use of switchable azobenzene attachment.

Appendix A

Copyright Notices

Figure 1.1 reprinted with permission from ACS Chemical Biology 3:38-50. Copyright (2008) American Chemical Society.

Figure 1.4a reprinted with permission from Journal of Organic Chemistry 72(18):6635-52. Copyright (2007) American Chemical Society.

Figure 1.4b reprinted by permission from Macmillan Publishers Ltd: Nature 369(6476), copyright 1994.

Figure 1.4c reproduced from Chemical Society Reviews 35(11):1043-55 (2006) with permission of The Royal Society of Chemistry.

Figure 1.4d Copyright 2014 American Chemical Society

Figure 1.4e Copyright 2014 American Chemical Society

Figure 1.5 reprinted by permission from Macmillan Publishers Ltd: Nature 440(7082):297-302, copyright (2006).

Figure 1.16a taken from the Journal of Immunology 183(11):726877 (2009). Copyright 2009. The American Association of Immunologists, Inc.

Figure 1.16b taken from Science 340:595-9 (2013). Reprinted with permission from AAAS.

Appendix B

Abbreviations

B.1 Abbreviations

ADP	Adenosine diphosphate
AFM	Atomic Force Microscopy
ATP	Adenosine triphosphate
AUC	Analytical Ultra Centrifugation
BLAST	Basic Local Alignment Search Tool
bZIP	Basic leucine zipper (protein domain)
CAP	Catabolite Activator Protein
CCHMM	Coiled-Coil domains with Hidden Markov Models
CCHMM-PROF	Profile-based version of CCHMM
CD	Circular Dichroism
CG	Coarse-Grained (simulations)
COMP	Cartilage Oligomeric Matrix Protein
Da	Dalton (atomic mass unit, 1g/mol)
DCM	Dichloromethane
DIPEA	N,N-diisopropylethylamine
DLS	Dynamic Light Scattering
DMF	N,N-dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DtxR	Diphtheria toxin repressor
Fmoc	9H-fluoren-9-ylmethoxycarbonyl

FRET	Förster/Fluorescence Resonance Energy Transfer
GROMACS	GROningen MAchine for Chemical Simulations
HIV	Human Immunodeficiency Virus
HPLC	High Performance Liquid Chromatography
KIH	Knobs-Into-Holes
LC	Liquid Chromatography
L-J	Lennard-Jones (potential)
MALDI	Matrix-Assisted Laser Desorption/Ionisation
MARCOIL	Hidden MARkov model-based program that predicts potential coiled-COIL domains in protein sequences
MBHA	Methylbenzhydryl amine
MD	Molecular Dynamics
MeCN	Acetonitrile
MetJ	Methionine repressor
MRE	Mean Residue Ellipticity
MW	Molecular Weight
NCL	Native Chemical Ligation
NMP	N-methylpyrrolidone
NMR	Nuclear Magnetic Resonance
OD	Optical Depth
P_i	Inorganic phosphate molecule
PBS	Phosphate Buffered Saline
PCOILS	Profile-based version of COILS
Pd(II)	Palladium (oxidation state 2; charge +2)
PDB	Protein Data Bank
PDMS	Polydimethylsiloxane
PSIBLAST	Position-Specific Iterated BLAST
Pt(II)	Platinum (oxidation state 2; charge +2)
PurR	Pur operon repressor
PyBOP	Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
rpm	revolutions per minute
RNA	Ribonucleic acid
ROI	Region of interest

SARS	Severe Acute Respiratory Syndrome
SE	Sedimentation Equilibrium
SV	Sedimentation Velocity
SPPS	Solid Phase Peptide Synthesis
TCEP-HCl	(tris(2-carboxyethyl)phosphine) hydrochloride
TFA	Trifluoroacetic acid
TIPS	Triisopropylsilane
ToF	Time-of-flight
Tris-HCl	Tris(hydroxymethyl)aminomethane hydrochloride
TrpR	Tryptophan repressor
UV	Ultraviolet
VMD	Visual Molecular Dynamics

Collaborators

ACM	Asahi Cano-Marques
EHCB	Elizabeth H. C. Bromley
MJZ	Martin J. Zuckermann
RBS	Richard B. Sessions

B.2 Amino Acid Abbreviations

Table B.1: Amino Acid One- and Three-letter Codes

Amino Acid Name	Three-letter code	One-letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamic Acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Appendix C

Product Masses

Table C.1: Peptide disulphide pairings and resultant masses

Peptide	Sequence	Calculated Mass
6-1	NKIAALKYKIAALKQEIAALEQGNNGC CGNGNEIAALEKKIAALKQENAALEQEIAALEY	6470
3-2	EIAALEQEIYALEQKNAALKKEIAALEQGNNGC CGNGNKIAALKQEIYALEQKNAALKQKIAALK	7096
4-5	NKIAALKQKIAQLKQENAALEQKIYALKQGNNGC CGNGNKIKALKQEIAALEYEINALEQ	6682
1-2	(CGNGNEIAALEKKIAALKQENAALEQEIAALEY)[1] CGNGNKIAALKQEIYALEQKNAALKQKIAALK	7039
3-4	EIAALEQEIYALEQKNAALKKEIAALEQGNNGC (NKIAALKQKIAQLKQENAALEQKIYALKQGNNGC)[1]	7395
6-5	NKIAALKYKIAALKQEIAALEQGNNGC CGNGNKIKALKQEIAALEYEINALEQ	5814

[1] All of the sequences are written from N to C termini. The brackets denote that the peptide is disulphide-bonded C-N, due to the position of its cysteine residue. The N-termini of peptides 1 and 2 bind, while the C termini of peptides 3 and 4 bind. This is clarified in Figure 3.12.

Table C.2: Peptide Masses

Name	Sequence (All written N-C)	No. of residues	Termini Modifications	Calculated Mass	Molecular Formula
1	CGNGNEIAALEKKIAALKQENA-ALEQEIAALEY	33	N-acetylated, C-amidated	3558.00	$C_{154}H_{254}N_{42}O_{52}S_1$
2	CGNGNKIAALKQEIYALEQKNA-ALKQKIAALK	32	N-acetylated, C-amidated	3483.11	$C_{154}H_{265}N_{45}O_{44}S_1$
3	EIAALEQEIYALEQKNAALKKE-IAALEQGNNGC	33	N-acetylated, C-amidated	3615.05	$C_{156}H_{257}N_{43}O_{53}S_1$
4	NKIAALKQKIAQLKQENAALEQ-KIYALKQGNNGC	34	N-acetylated, C-amidated	3782.39	$C_{165}H_{282}N_{50}O_{49}S_1$
5	CGNGNKIKALKQEIAALEYEIN-ALEQ	26	N-acetylated, C-amidated	2902.30	$C_{126}H_{209}N_{35}O_{41}S_1$
6	NKIAALKYKIAALKQEIAALEQ-GNNGC	27	N-acetylated, C-amidated	2914.40	$C_{128}H_{217}N_{37}O_{38}S_1$
ϵ	CGNGNEIAALEYEINALEQKIAALK	25	N-acetylated, C-amidated	2717.07	$C_{118}H_{194}N_{32}O_{39}S_1$
ϵ2	EIAALEYEINALEQKIAALKQKN-CAGAANACNKIAALKQEIYALEQ-KNAALKQKIAALK	59	C-amidated	6354.43	$C_{281}H_{477}N_{79}O_{83}S_2$

Table C.2 continued

Name	Sequence (All written N-C)	No. of residues	Termini Modifications	Calculated Mass	Molecular Formula
$\epsilon 2$ CL	EIAALEYEINALEQKIAALK-QKNCAGAANACNKIAALKQEIYA-LEQKNAALKQKIAALK	59	C-amidated, azobenzene bound between cysteines	6642	$C_{297}H_{489}N_{83}O_{85}S_2$

Appendix D

Experimental Methods

D.1 Peptide Synthesis

All peptides were synthesised using a CEM Liberty 1 Automated Microwave Peptide Synthesiser, and PepDriver software, on a 0.1mM scale. Table C.2 lists all of the peptide sequences synthesised and used in this thesis, with modifications and masses. Peptides 1 and 6 were provided in crude form by Marc Bruning (Woolfson Lab, Bristol University). Piperidine (20% in DMF) was the deprotect solution, PyBOP (0.5M in DMF) the activator, and 2M DIPEA in NMP the activator base for all syntheses. Fmoc amino acids (dissolved in DMF) were used at a concentration of 0.2M, and all Fmoc groups were removed in the synthesis process. Rink amide resin, which has an NH_2 group, was used for all syntheses, so upon removal from the resin, peptides were amidated at the C-terminus. The following amino acids were used (side chain protecting groups shown in brackets): Fmoc-Gly-OH, Fmoc-L-Tyr(tBu)-OH.H_2O, Fmoc-L-Glu(tBu)-OH, Fmoc-L-Leu-OH, Fmoc-Asn(Trt)-OH, Fmoc-L-Gln(Trt)-OH, Fmoc-L-Ala-OH.H_2O, Fmoc-L-Lys(Boc)-OH, Fmoc-L-Ile-OH, Fmoc-L-Cys(Trt)-OH. All of the amino acids used were purchased as powders from AGTC Bioproducts Ltd, along with PyBOP, and solutions prepared on the day of a synthesis. DIPEA, NMP, and Piperidine were all purchased from Sigma Aldrich. DMF was purchased from Rathburn Chemicals and AGTC Bioproducts. Rink Amide MBHA resin, 100-200 mesh, substitution 0.59mmole/g, was purchased from Novabiochem. Peptides 1 to 6 were all synthesised using the standard method described below (Section D.1.1). Deviation from this method for $\epsilon 2$ is discussed in Section D.1.2.

D.1.1 Standard Synthesis

1. The required peptide sequence was put into the PepDriver software, and all of the solutions made up for a synthesis. Any solutions remaining from previous syntheses (unless syntheses were performed directly one after another) were replaced with fresh solutions. The piperidine solution could be kept for much longer.

2. Resin was weighed out and swelled in a small amount of DMF (∼1ml) in the reaction vessel prior to synthesis. This was then replaced in the centre of the microwave, ready for synthesis. Additionally, the machine automatically allowed 15 minutes for swelling as part of its run sequence.

3. The automated synthesis was then carried out. The synthesiser takes around 19 hours to synthesise each of the hub peptides discussed in this report (varying with length/composition) and hence syntheses were usually carried out overnight. Once the synthesis was completed, the peptide, still attached to the resin, was removed from the machine.

4. The peptide resin beads were transferred to filter tubes (with open tops, and filters and stoppers at the bottom), and acetylated using acetic anhydride (20% in DMF, 5ml total volume per 0.1mM scale peptide synthesis), stirred for 30 minutes at room temperature using a magnetic stirrer and stirring plate.

5. The acetic anhydride-DMF solution was removed from the peptide resin via the stopper, and the resin rinsed several times with DMF, before it was dried using a pump. (In later syntheses, they were further rinsed and shrunk with diethylether, as cleavage does not require swollen resin).

6. Once the peptide resin was dry, the stopper was replaced, and the peptide cleaved from the resin. This was done using a solution of TFA (95%), TIPS (2.5%) and distilled water (2.5%). If cysteine was present, as it was in the hub peptides, 1,2-ethanedithiol was also added, in the same volume as the TIPS and distilled water. The resin (with peptide attached) was left in this solution for at least three hours, stirred constantly using a magnetic stirrer and stirring plate. If the resin turned red immediately after addition of the solution, more TIPS was added. After one hour of stirring, the colour was again checked to ensure that sufficient scavenger was present to prevent unwanted reactions between the peptide and the removed groups.

7. After three hours, the peptide should be in solution, and separable from the resin via the filter. The peptide was precipitated into chilled diethylether via the tube filter. The tubes of peptide in cleavage solution and ether were spun down in a centrifuge at 3000 rpm for 10 minutes in order to separate the cleavage chemicals from the peptide. The peptide settles to the bottom of the tubes (the solvent remains cloudy, and more centrifugation is needed, if this does not occur) and the solvents are then poured off carefully, leaving the white powder sedimented at the bottom of the tubes. More diethylether was then added to each tube, the powder re-dispersed and centrifugation repeated in order to remove any remaining cleavage solution. The solvent was again poured off, leaving the peptide.

8. The peptides were then dissolved in 5ml of distilled water per tube, frozen and freeze dried, after which they were kept in a freezer for significant amounts of time.

D.1.2 Non-standard Syntheses

ϵ2 Synthesis

ϵ2 was synthesised by Asahi Cano-Marques. The same chemicals were used as previously stated.

There were some changes to the standard synthesis scheme:

- The peptide was synthesised in three stages, with test cleaves carried out after synthesis of 30, 45 and 59 residues.
- The resin swelling time was longer - two hours in DCM prior to synthesis, six hours in DCM prior to addition of residues 30 to 15, and 16 hours in DMF prior to addition of residues 14 to 1.
- Resin was shrunk in diethylether prior to storage between stages and prior to cleavage, to remove DMF.
- Fmoc deprotection was carried out using two treatments of piperidine in DMF, for 5 and 10 minutes respectively.
- Most of the couplings were double couplings (residues 44-30, 28-15, 12-2), while residues 59, 29, 14, 13 and 1 were triple coupled, and 58-45 single coupled. Single couplings used a fivefold excess of amino acids, 10 equivalents of PyBOP, and 20

equivalents of DIPEA. Double couplings used single coupling amounts of amino acid, DIPEA and PyBOP, reacted, removed and repeated, so double couplings used twice the single coupling quantities in total.

- Reaction volumes were also varied - residues 58-30, 28-15 and 12-2 used double volumes. These volumes were at the same concentrations as previously stated, and hence contained twice the number of equivalents. For double volumes in the first stage, this meant a total of double the equivalents of the single couplings above, and quadruple equivalents for those with double couplings. Due to the resin being split after residue 30 (half was saved for later additions to be carried out via an alternative method if the initial synthesis scheme failed), these equivalents were in fact doubled again for double volumes in stages 2 and 3, as the additions were left at previous volumes/concentrations, even though only half the amount of resin remained.

Once $\epsilon 2$ was synthesised, it was purified, and some of it was crosslinked with azobenzene via its cysteine residues:

1. A 1.1ml, 1.4mM solution of the peptide in water was stirred at room temperature with 5ml, 5mM of Tris-HCl (pH 8.5) and 5ml, 5mM TCEP-HCl (pH 8.5) for at least 2 hours under N_2, then covered by aluminium foil to protect from the light, and heated to 40° C.

2. Once the solution was at 40° C, 2ml of 1.1mM azobenzene in DMSO was added dropwise to the solution, and stirred for 20 minutes. This was followed by 2ml of 18mM azobenzene in DMSO (added dropwise) and a further 20 minute stirring.

3. The solution was then exposed to the light and stirred for a final 20 minutes.

4. The product was then lyophilised and purified using HPLC.

D.2 MALDI-ToF

The MALDI-ToF experiments were done through the Chemistry Department Mass Spectrometry Service, using a Bruker Daltonics Autoflex II ToF/ToF machine. The peptides were mixed with α-cyano-4-hydroxycinnamic acid matrix.

D.3 HPLC

All peptides were purified using a C_{18} column with a Perkin Elmer 785A UV/Visible Detector, and Series 200 LC pump, and Clarity software. Peptides were dissolved in water, or, if not fully soluble in water, were wet with a small amount of acetonitrile, followed by water. Samples were spun down, and also filtered in some cases, in order to remove any particulate matter from the sample before injection into the HPLC system. HPLC was carried out using gradients from Solution A (95% distilled water, 5% acetonitrile, +0.1% TFA) to Solution B (95% acetonitrile, 5% water, +0.1% TFA), and monitoring elution of the Tyrosine-containing peptides via the 280nm wavelength.
An example of a HPLC method used:

- 3 minutes at 100% solution A
- 40 minute linear gradient from 100% solution A to 100% solution B
- 10 minutes at 100% solution B
- 10 minute linear gradient from 100% solution B to 100% solution A
- 5 minutes at 100% solution A

Fractions were collected as they left the machine and peptide fractions identified using MALDI-ToF. Methods were lengthened and gradients modified in some cases to try to increase resolution of peaks, and better separate products.

D.4 Disulphide Bonding

0.017g of Aldrithiol-2 was dissolved in 1ml of acetonitrile (MeCN). Each of the five peptides to be reacted with Aldrithiol-2 were dissolved in water, to a concentration of ∼1mg/ml, and 40μl of Aldrithiol in MeCN was added per millilitre of peptide solution, providing a tenfold excess of Aldrithiol-2. Each solution was then agitated and freeze-dried. HPLC purification was then used to separate the peptides from excess Aldrithiol-2, and side products from the reaction. Each aldrithiol peptide and its partner peptide were then reacted together in water, and the disulphide-bonded pair separated from other products via HPLC purification.

D.5 Samples for Characterisation

For each peptide, the purest fractions from HPLC were combined and dissolved in water, and these were kept as stock solutions. The stock solutions were made to approximately 1mg/ml concentrations and more accurate values were then found using UV-Vis spectroscopy. These stock solutions were then used for CD and DLS characterisation. CD and DLS measurements were carried out in phosphate buffer solution.

D.6 UV-Vis Measurements

D.6.1 Concentration Determination

Each of the six regular hub peptides contain one tyrosine residue, and it was the UV absorption of this which was used to measure the concentration of samples. Using the tyrosine absorption to measure tyrosine concentration gives peptide concentration, as they exist in a 1:1 ratio. For $\epsilon 2$, which has two tyrosines, there is a trivial change in the concentration calculation (Appendix F.1). All UV-Vis spectra for non-fluorescent peptides were carried out between 200 and 320nm or 340nm, on a UNICAM UV2 UV-Vis spectrometer, using a reduced volume quartz cuvette of 1cm path length. Most measurements were carried out on peptides dissolved in water. If there were signs of unusual activity, such as aggregation, PBS buffer was used. Measurements involved taking three spectra of each sample, and taking measurements of three samples for each peptide, allowing an average to be calculated.

UV Measurement Protocol:

1. Peptide samples of 10% stock concentration were prepared, i.e. 30μl of stock solution, and 270μl water.

2. A baseline was carried out using the UV cuvette filled with distilled water.

3. The sample was measured three times, and files converted to ASCII. The 280nm peak should have an OD approximately between 0.1 and 1. If solutions were found to be too weak/strong to make meaningful measurements, different concentrations were prepared, to achieve absorptions in the 0.1-1.0 region.

4. Once all measurements were taken, the concentrations were determined using the absorption at 280nm, compared with the baseline (Equations in Appendix F.1).

D.6.2 Switching Effects

For ϵ2CL, a UV spectrum was taken of the sample prepared for CD (using a 0.1cm cuvette, and 0.9cm spacer). After irradiation and CD measurements, the sample was measured again, and at intervals after this, to monitor the spectrum as the system returned towards its pre-irradiated state.

D.7 Circular Dichroism

CD measurements were taken using a JASCO J1500 (earlier measurements with a J-810) spectropolarimeter with a 150W Xenon arc lamp light source, and J-815 MCB-100 mini circulation bath and peltier stage to control temperature. The machine is purged of oxygen with nitrogen gas before and during use, as oxygen absorbs strongly in the region of interest, and to prevent ozone production.

1. 300μl, 20μM solutions were made, with the required volume of peptide solution from the stock solutions, 10% (30μl) PBS (10x concentrated), and the remainder of the 300μl, distilled water.

2. Spectra were collected at 20 °C over a wavelength (λ) range of 260-190nm (other parameters shown in Table D.1).

3. Melts were also performed (parameters shown in Tables D.2 and D.3).

4. The results were then converted into mean residue ellipticity units, using Equation F.12.

A typical PBS solution baseline spectrum (formed from multiple mesurements of PBS solutions at various points between the peptide measurements) was removed from all of the peptide λ vs. CD spectra. I took five λ vs. CD spectra at 20° C both pre- and post-melt for each solution and one melt.
For mixtures, the concentration of each peptide was 20μM, so the required volume of each peptide was added to 30μl of PBS (10x) and the water required to make a 300μl volume sample.

D.7.1 Concentration of ϵ2CL

The azobenzene moiety of ϵ2CL meant that the UV signal from Tyrosine could not be seen, and hence comparison of the CD of ϵ2CL with ϵ2 was used to estimate the concentration

Table D.1: Parameters used in CD measurements over a range of wavelengths at 20 °C

Parameter	**Explanation/Options**	**Value Used**
Sensitivity	'Low', 'Standard' or 'High'	Standard
Wavelength	Wavelength range used for measurements	260-190nm
Data Pitch	Interval between data points recorded	1nm
Mode	'Step' or 'Continuous'	Continuous
Scan speed	Wavelengths scanned per interval time	50nm/min
Response	Time over which data is averaged (smooths out instabilities), related to scan speed	2s
Bandwidth	Similar to resolution; can be coarse as the peaks measured here are broad	1nm
Accumulation	No. of scans averaged to get result	1[1]

[1] Instead of using accumulation, multiple scans were taken manually and results averaged

Table D.2: Parameters used in CD melt experiments - wavelength options

Parameter	**Explanation/Options**	**Value Used**
Sensitivity	'Low', 'Standard' or 'High'	Standard
Wavelength	Wavelength range used for measurements	260-190nm
Data Pitch	Interval between data points recorded	1nm
Mode	'Step' or 'Continuous'	Continuous
Scan speed	Wavelengths scanned per interval time	50nm/min
Response	Time over which data is averaged (smooths out instabilities), related to scan speed	2s
Bandwidth	Similar to resolution; can be coarse as the peaks measured here are broad	1nm
Accumulation	No. of scans averaged to get result	1[1]
Delay time	Time machine waits between achieving required starting temperature and starting measurement	30s

[1] Instead of using accumulation, multiple scans were taken manually and results averaged

Table D.3: Parameters used in CD melt experiments - temperature options

Parameter	**Explanation/Options**	**Value Used**
Wavelength	Wavelength CD is measured at as a function of temperature	222nm
Temperature	Temperature range measured over	10-90 °C
Data Pitch	Interval between data points recorded	1 °C
Delay time	Time machine waits between achieving required starting temperature (10 °C) and starting measurement	30s
Scan speed	Temperature range covered per interval time	60 °C/hour
Sensitivity	Low, standard, high	Standard
Response	Time over which data is averaged (smooths out instabilities), related to scan speed	8s
Bandwidth	Similar to resolution; can be coarse as the peaks measured here are broad	3nm

of ϵ2CL.

D.8 DLS

DLS measurements were taken on a Malvern Zetasizer μV machine using Sarstedt disposable transparent cuvettes. 100μM samples were made in PBS buffer solution. For multiple-peptide samples, each peptide was present at a concentration of 100μM.

D.9 Glutathione Experiments

Concentrations of the 1-6, 2-3 and 4-5 disulphide bonded peptides (dissolved in water) were measured using a Thermo Scientific Nanodrop 1000, so that a 1:1:1 ratio could be produced as accurately as possible using low concentrations and low volumes of peptides. In order to encourage disulphide exchange, the peptides were added to PBS buffer solution containing a low concentration of reduced glutathione. As the exchange was expected to be rapid, the solution was left for under 10 minutes, after which the volume of the sample was almost doubled (from 58 to 108 μl) with TFA in water (4μl in 15ml water), to lower the

pH and quench the reaction. This sample was then submitted for MALDI-ToF analysis. A 1:1:1 ratio of the three disulphide bonded peptides was also added into a separate 50μl TFA in water sample for submission for MALDI.

Appendix E

Spectra

E.1 MALDI Spectra

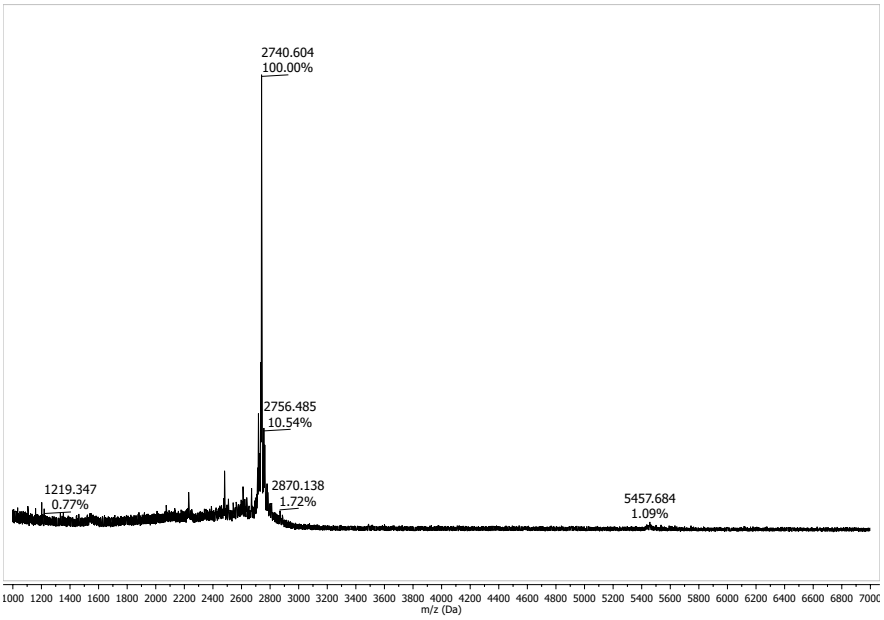
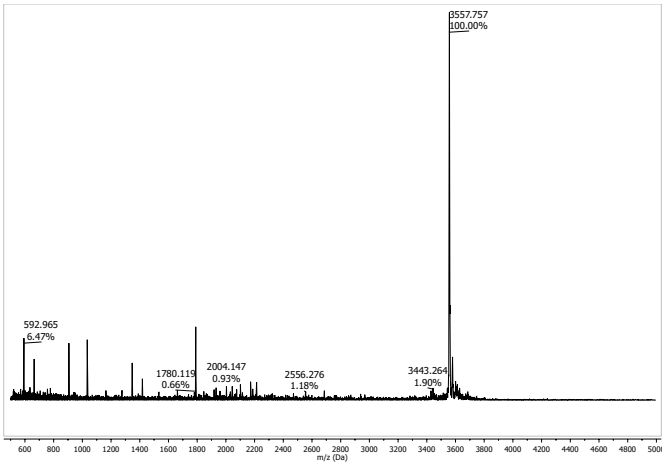
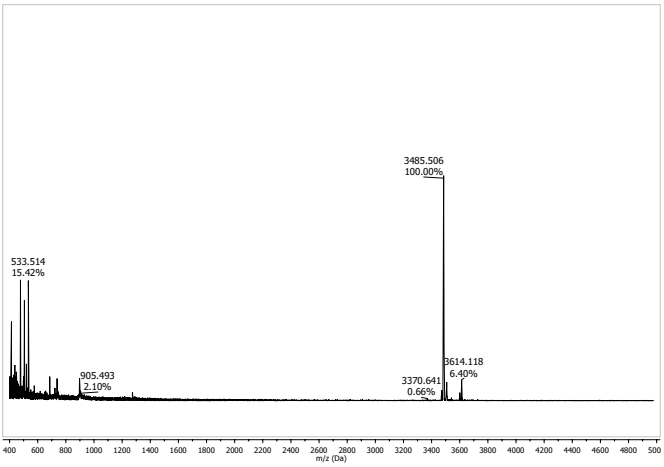


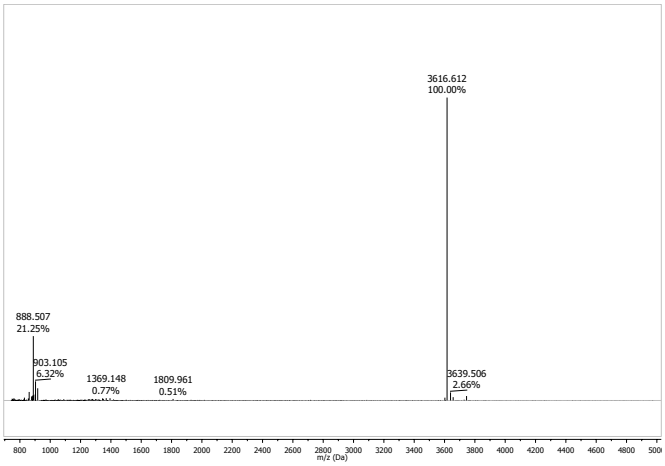
Figure E.1: MALDI-ToF spectrum of ϵ (synthesised by EHCB).



(a) Peptide 1

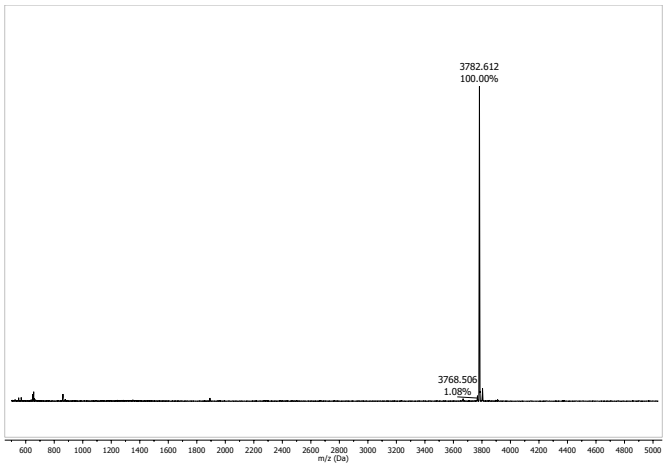


(b) Peptide 2

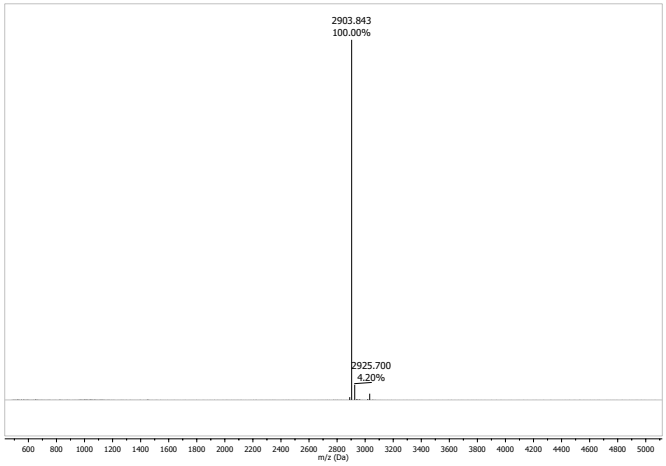


(c) Peptide 3

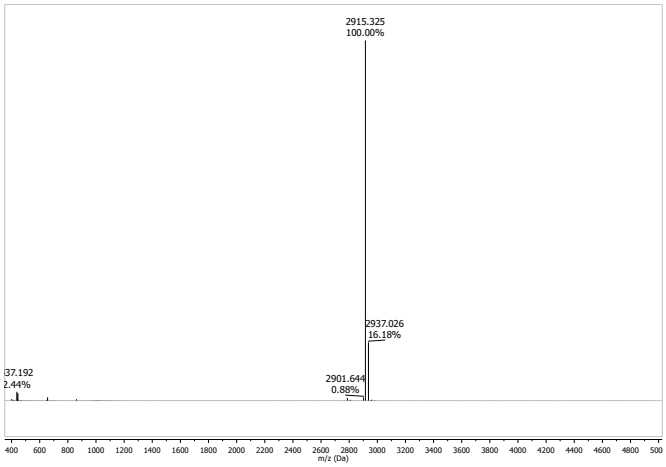
Figure E.2: MALDI-ToF spectra of the individual hub peptides 1, 2 and 3.



(d) Peptide 4

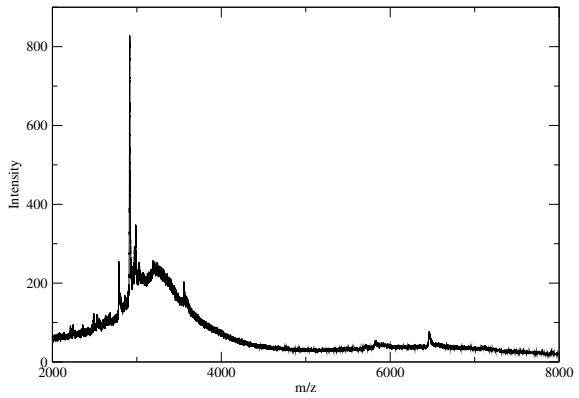


(e) Peptide 5

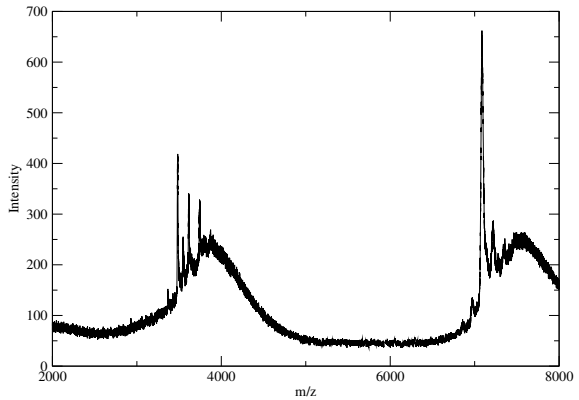


(f) Peptide 6

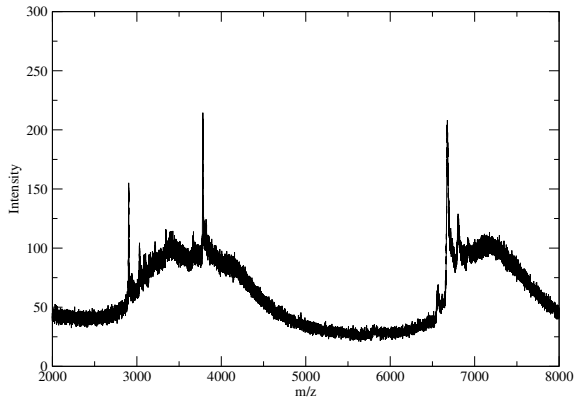
Figure E.2: MALDI-ToF spectra of the individual hub peptides 4, 5 and 6.



(a) 1-6



(b) 2-3



(c) 4-5

Figure E.3: MALDI-ToF spectra of the disulphide-bonded hub peptides. The main peak at higher masses is the desired product in each, with masses of 6461, 7086, and 6677 Da measured respectively.

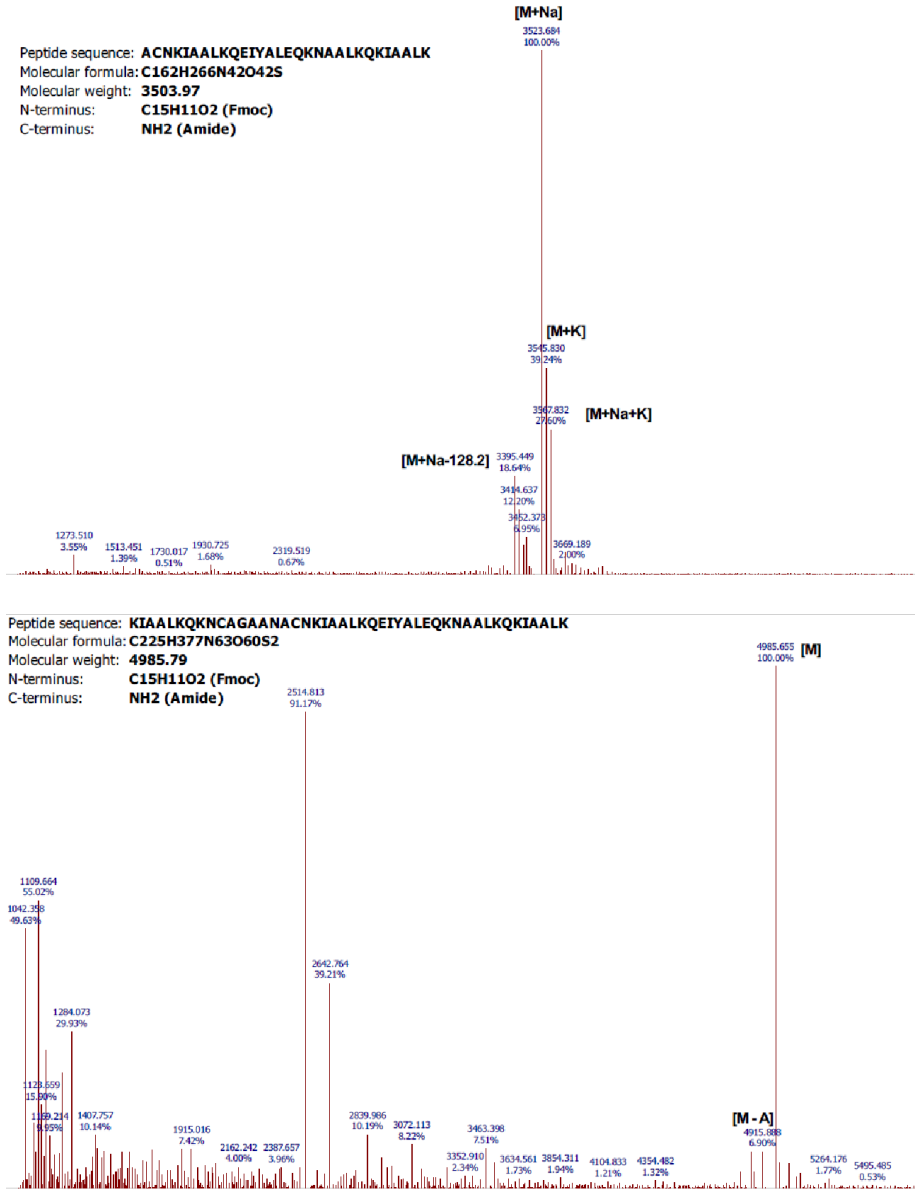


Figure E.4: MALDI-ToF spectra of ϵ2 during synthesis, after a) 30 residues and b) 45 residues (credit: ACM).

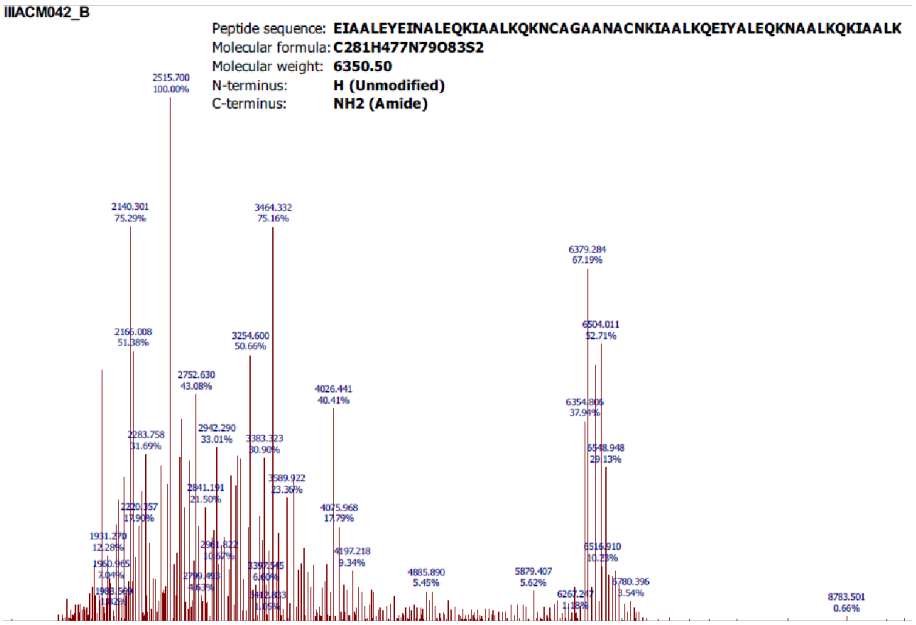


Figure E.5: Final MALDI-ToF spectrum of crude ϵ2. The spectrum does not look ideal, but there are clearly several peaks near the peptide mass, indicating some peptide and some near-correct peptide constituents (credit: ACM).

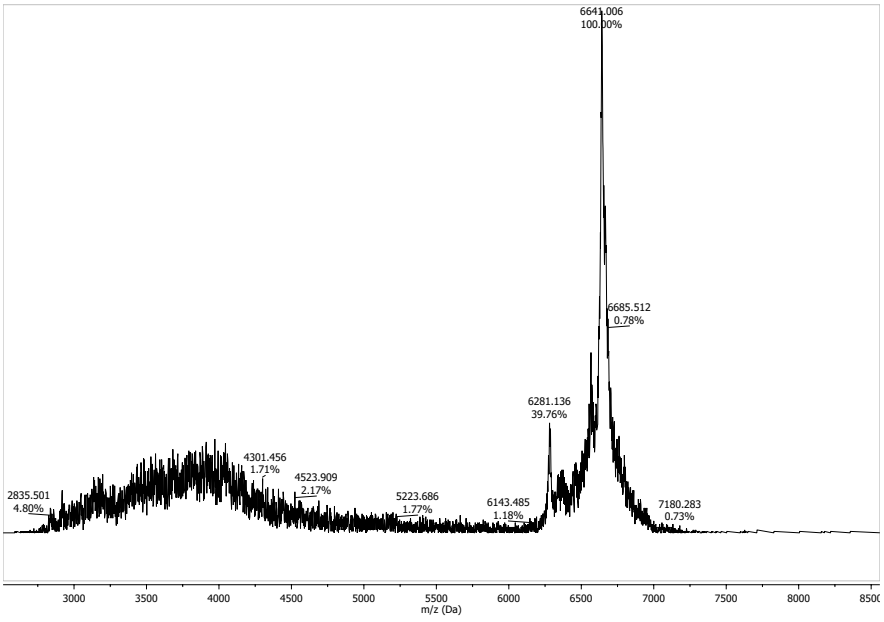


Figure E.6: MALDI-ToF spectrum of ϵ2CL. The peak is at the desired mass of 6642 Da. ϵ2CL was synthesised by ACM.

E.2 CD Melt Data for Peptide Mixtures

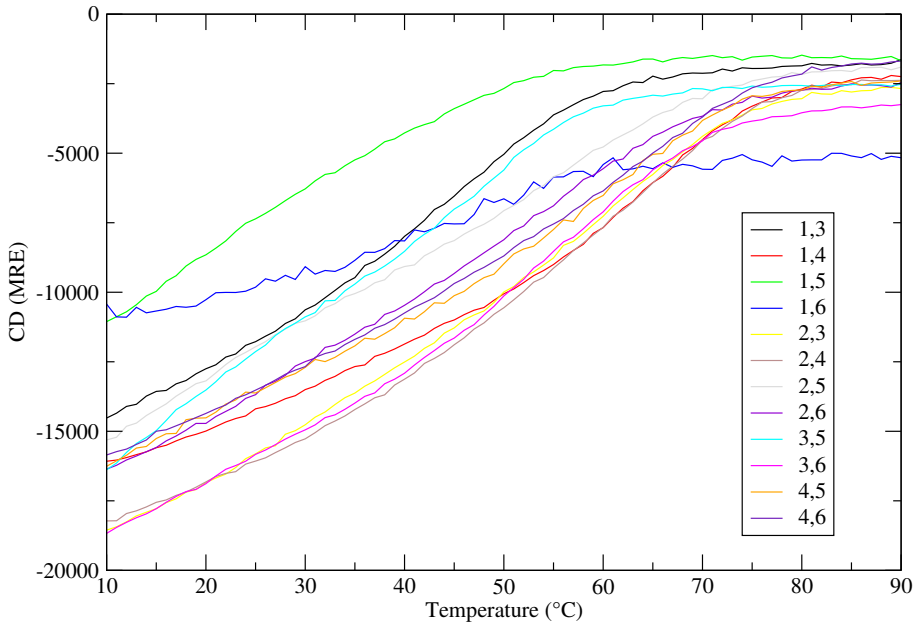


Figure E.7: Melt spectra for the 12 non-designed two-peptide mixtures, taken at 20μM concentration per peptide in PBS buffer (pH 7.4).

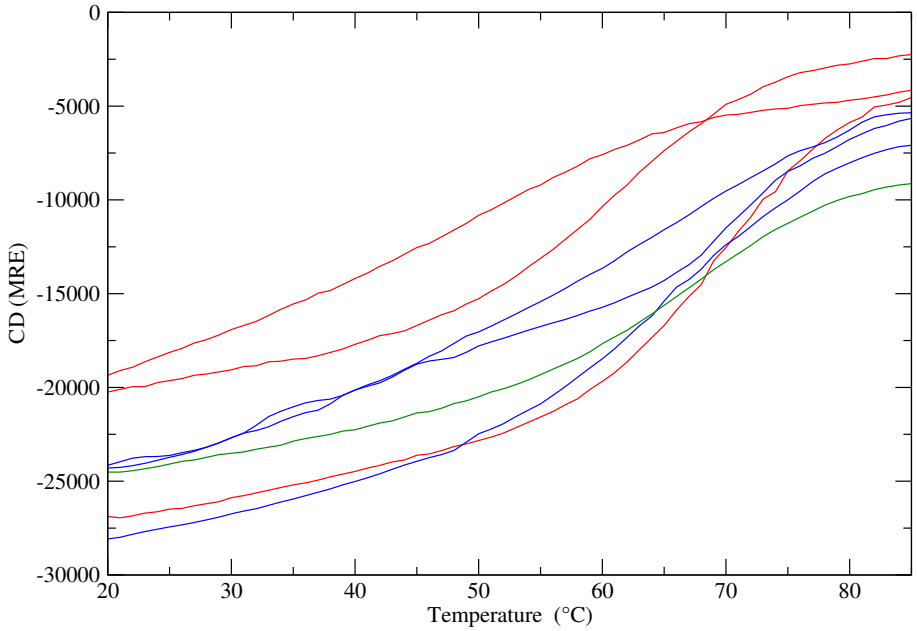
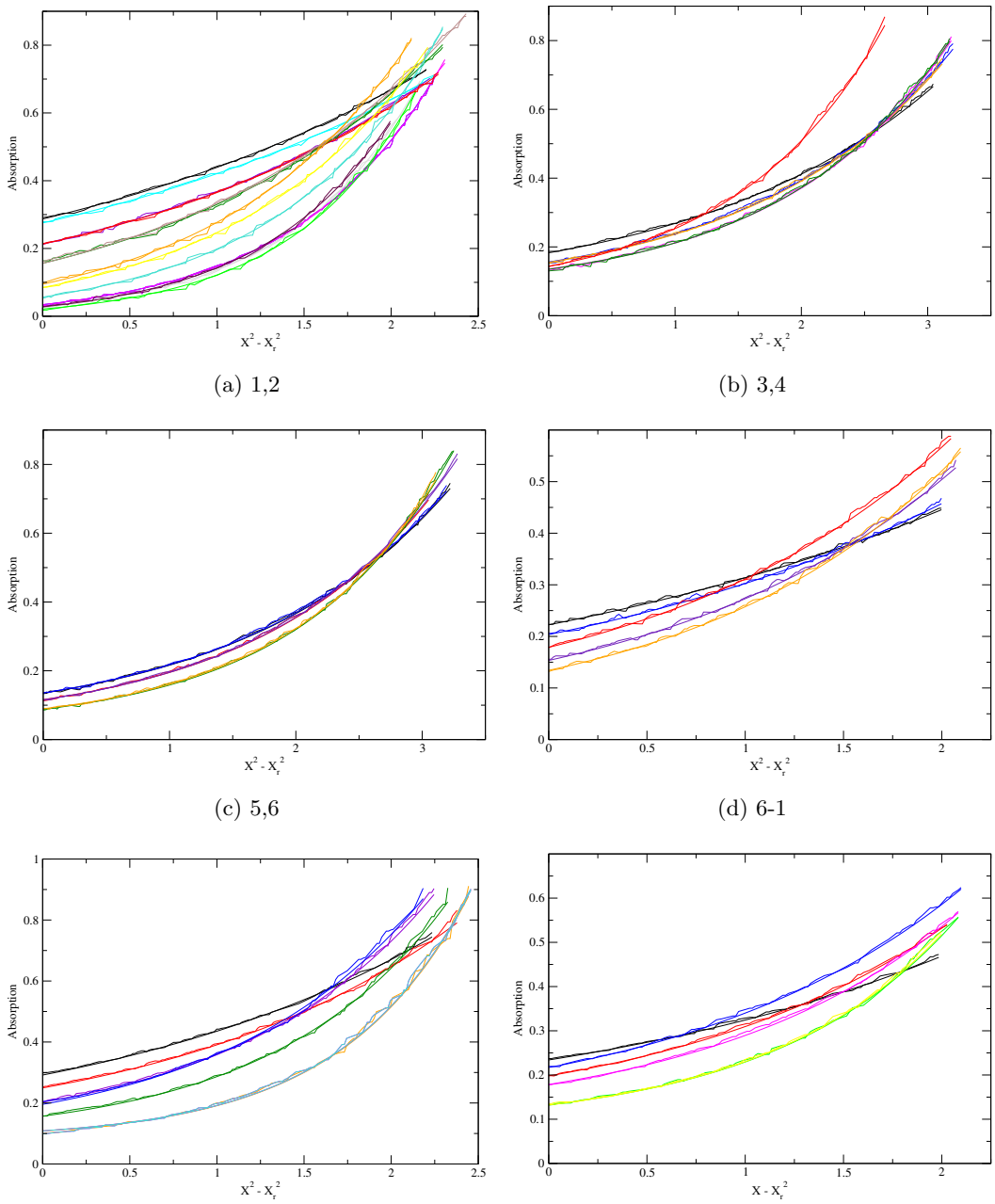


Figure E.8: Melt data of disulphide-bonded peptides, taken by EHCB. Red lines are the spectra of single disulphide peptides, pairs are shown in blue, while the full hub is shown in green.

E.3 Raw AUC Data

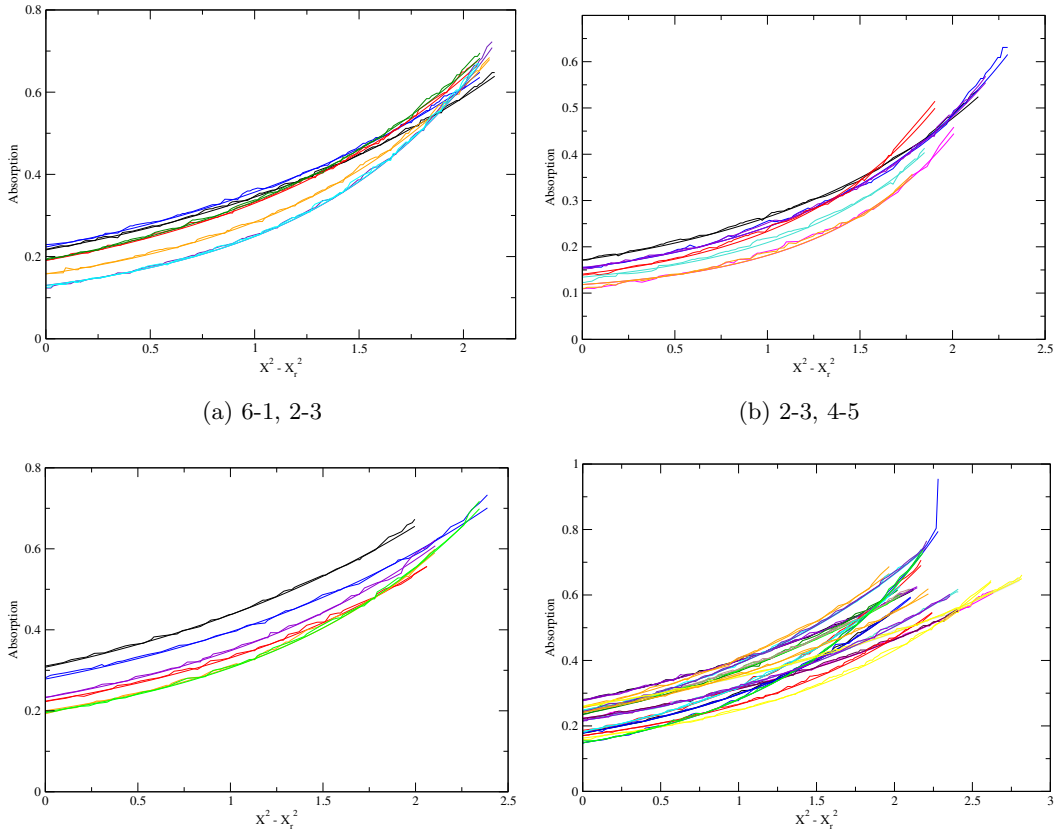


(a) 1,2 (b) 3,4

(c) 5,6 (d) 6-1

(e) 2-3 (f) 4-5

Figure E.9: AUC data of coiled-coil pairs and disulphide-bonded peptides.



(a) 6-1, 2-3

(b) 2-3, 4-5

(c) 4-5, 6-1

(d) 6-1, 2-3, 4-5

Figure E.10: AUC data of pairs of disulphide-bonded peptides, and all three disulphide-bonded peptides together.

Appendix F

UV and CD Calculations

F.1 Calculation of Concentration from UV Absorbance

The Beer-Lambert law:

$$A = \epsilon c l \tag{F.1}$$

where A is absorbance (measured), c is concentration (desired), l is path length (known) and ϵ is extinction coefficient (known from literature). This gives:

$$c_{\text{sample}} = \frac{A}{\epsilon \cdot l} \tag{F.2}$$

For these samples:

$$c_{\text{sample}}(M) = \frac{A}{\epsilon(M^{-1}cm^{-1}) \cdot 1(cm)} \tag{F.3}$$

where ϵ depends on the number of tyrosine residues present:

$$\epsilon_{\text{Tyr}} = 1280 M^{-1} cm^{-1} \tag{F.4}$$

$$\epsilon_{\text{2Tyr}} = 2 \cdot 1280 M^{-1} cm^{-1} = 2560 M^{-1} cm^{-1} \tag{F.5}$$

Once the concentration of the sample has been calculated, the stock concentration is found:

$$c_{\text{stock}}(M) = c_{\text{sample}} \cdot d \tag{F.6}$$

where d is the dilution factor

$$d = \frac{V_{\text{total sample}}(\mu l)}{V_{\text{stock solution in sample}}(\mu l)} \tag{F.7}$$

A is measured directly by the spectrometer, but there is a baseline, a non-tyrosine contribution, which needs to be removed. Therefore

$$A = A_{\text{peak}} - A_{\text{baseline}} \tag{F.8}$$

A_{baseline} was found by extrapolating a baseline from the best fit to the data at 310-320 nm, which was normally (and this was confirmed by eye) unaffected by the peak.

F.2 Circular Dichroism Units, Conversions and Calculations

F.2.1 Explanation of CD Units

Circular dichroism is recorded by the spectropolarimeter in mdeg, units of ellipticity. This is usually then converted to molar ellipticity, M_θ, or mean residue ellipticity (MRE), which is molar ellipticity per peptide bond.

$$\begin{aligned} M_\theta &= \frac{deg \cdot cm^2}{dmol} \\ M_\theta &= \frac{deg \cdot cm^3}{dmol \cdot cm} \\ M_\theta &= \frac{10deg \cdot cm^3}{mol \cdot cm} \\ M_\theta &= \frac{1000deg \cdot cm^3}{100mol \cdot cm} \\ \frac{1000cm^3}{mol} &= M^{-1} \\ 100cm &= 1m \\ M_\theta &= \frac{degM^{-1}}{100 \cdot cm} \\ M_\theta &= \frac{degM^{-1}}{m} \\ M_\theta &= \deg M^{-1}m^{-1} \end{aligned} \tag{F.9}$$

This molar ellipticity, M_θ, is historically displayed in the units degcm^{2}dmol^{-1}, which is equivalent to deg M^{-1}m^{-1} (as shown in Equation F.9). It is therefore common to use Equation F.10, with pathlength in cm, and a balancing factor of 100 to convert to metres. This is then equivalent to the traditional units. The units of Mean Residue Ellipticity (MRE) are degcm^{2}dmol^{-1}res^{-1}.

$$M_\theta \left(\frac{degcm^2}{dmol} \right) = \frac{100\theta(deg)}{l(cm)M(moles/litre)} \tag{F.10}$$

F.2.2 Conversion of Raw Data to Final Units

In order to present the CD data in its final form, an average of the CD runs (usually five) at 20°C (pre-melt) was taken for each sample. A PBS baseline was removed from the sample average to give the corrected measurements in units of mdeg. The data was then converted

into molar ellipticity by dividing the corrected values by the molar concentration (other factors cancel in our case, Equation F.11). To then convert to MRE, the molar ellipticity is divided by peptide bond number (Equation F.12). Our peptides have one more peptide bond than residue, due to there being one less peptide bond than the number of residues between the residues, plus one bond for each of the acetylation and amidation of the N- and C- termini respectively. For the mixtures this calculation is slightly more complicated, and no molar ellipticity was calculated, only MRE, as shown in Equation F.13, where θ is again the baseline-corrected data.

$$\begin{aligned} M_\theta &= \frac{100\theta(deg)}{l(cm) \cdot M(moles/litre)} \\ l &= 0.1cm \\ M_\theta &= \frac{100\frac{\theta(mdeg)}{1000}}{0.1 \cdot M(moles/litre)} \\ M_\theta &= \frac{\theta(mdeg)}{M(moles/litre)} \end{aligned} \tag{F.11}$$

$$MRE = \frac{\theta(mdeg)}{M(moles/litre) \cdot N_{peptidebonds}} \tag{F.12}$$

$$MRE_{mixtureA,B} = \frac{\theta(mdeg)}{M_A(moles/litre) \cdot N_{\text{peptide bonds, A}} + M_B(moles/litre) \cdot N_{\text{peptide bonds, B}}} \tag{F.13}$$

F.2.3 Theoretical Curves for Mixtures

Theoretical spectra of mixtures were calculated by summing the baseline-corrected data of the two contributing peptides, and then converting to MRE in the same way as with the measured mixtures - dividing by the sum of the concentration-peptide bond product of each peptide (Equation F.14).

$$MRE_{\text{theor. A,B}} = \frac{\theta_A(mdeg) + \theta_B(mdeg)}{M_A(moles/litre) \cdot N_{\text{peptide bonds, A}} + M_B(moles/litre) \cdot N_{\text{peptide bonds, B}}} \tag{F.14}$$

Appendix G

Coarse-grained Simulations

Martin Zuckermann carried out 3D coarse-grained Langevin dynamics simulations, solving the overdamped Langevin equation (Equation G.1).

$$\Delta x_i^{(j)} = \frac{F_i^{(j)} \Delta t}{\gamma} + \left(\frac{2kT\Delta t}{\gamma} \right)^{1/2} \zeta_i^{(j)}(t) \tag{G.1}$$

$\Delta x_i^{(j)}$ is the change in the ith co-ordinate (i = 1,2,3) of the jth bar motor component (j = 1,2,3, representing F_A, H and F_B, the three spherical components, shown in Figure 4.2) over a time Δt, at time t. $F_i^{(j)}$ is the ith component of the force acting on the jth component at time t, which is a sum of an internal conservative force $\left(F_{\mathrm{Ci}}^{(\mathrm{j})}\right)$ and an external force $\left(F_{\mathrm{Ei}}^{(\mathrm{j})}\right)$. In these simulations, $F_{\mathrm{Ei}}^{(\mathrm{j})}$ is set to zero. $F_{\mathrm{Ci}}^{(\mathrm{j})}$ is the negative gradient of a potential, the sum of W_H, W_{L-J}, W_{SB} and W_{BE}. W_H is a harmonic potential defining length l. W_{L-J} is a Lennard-Jones repulsive interaction representing the excluded volume between the feet F_A and F_B. W_{SB} is the specific binding potential:

$$W_{\mathrm{SB}}(r_j) = -V_{\mathrm{SB}}\, e^{\frac{-r_j^2}{d_b^2}} \tag{G.2}$$

where r_j is the separation between component j and its nearest specific binding site on the track; d_b is the effective range of the specific binding interaction strength V_{SB}.
W_{BE}, the bending energy, fixes the angle θ_0 between the two stiff bonds (legs), if V_{BE} is positive and sufficiently large:

$$W_{\mathrm{BE}} = -V_{\mathrm{BE}}\, cos^2(\theta - \theta_0) \tag{G.3}$$

γ, the drag coefficient for each of the spheres is given by the Stokes-Einstein equation:

$$\gamma = 6\pi\eta r \tag{G.4}$$

where η is the viscosity of the solution, and r is the radius of the sphere.
$\zeta_i^{(j)}(t)$ is a random number from a zero-mean Gaussian distribution where

$$\left\langle \zeta_i^{(j)}(t) x \zeta_{i'}^{(j')}(t') \right\rangle = \delta_{ii'}\delta_{jj'}\delta(t-t') \tag{G.5}$$

For pulses where a ligand is absent, $\mathrm{W}_{SB}(\mathrm{r}_j)$ is set to zero.

It is through the binding energy equation that the two simulated scenarios were created. In scenario one, V_{BE} is always 8000kT, so the angle is fixed, and θ_0 is switched from 180° to 80° at the start of the pulse which switches from $\theta_0 = \psi$ to $\theta_0 = \phi$. In scenario two, θ_0 is kept at 180°, but in the $\theta_0 = \psi$ pulses, V_{BE} is 8000kT, and in the $\theta_0 = \phi$ pulses, V_{BE} = 0, allowing θ_0 to vary.

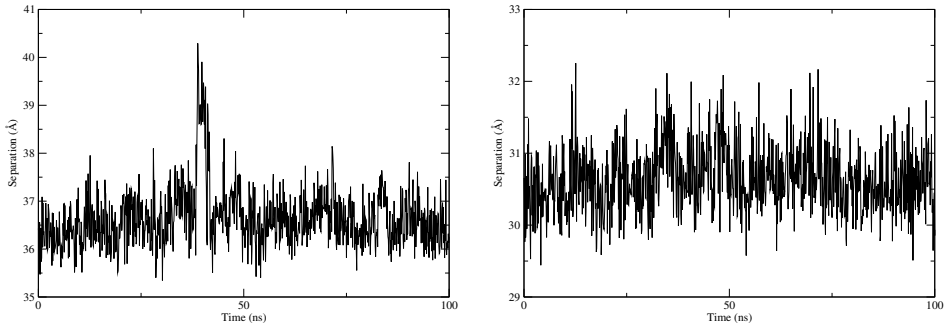
Appendix H

Molecular Dynamics

H.1 Residue Stability Checks For End-to-end Separation Measurements

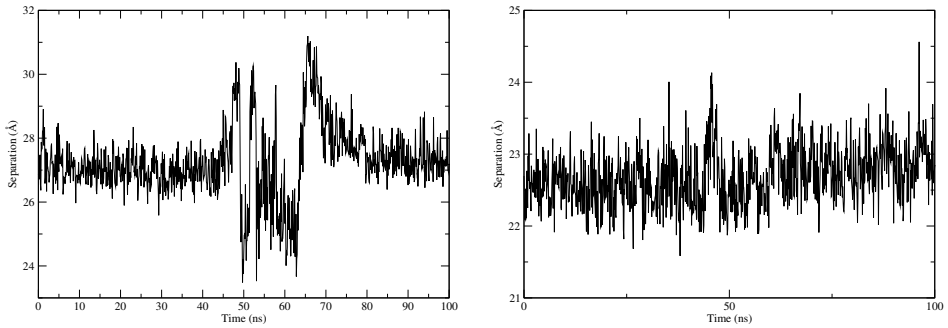
The evolution of the separations shown in Figure 4.11b were checked for all simulation runs to ensure that the residues chosen for the end-to-end separations remained part of their respective coiled coils for the duration of the simulation. The same type of measurements were taken for the residues chosen for vector angle calculations, shown in Figure 4.12c. Figure H.2 shows the evolution of the separation between six pairs of residues for the initial azobenzene-free system simulation. Four of those shown are those in Figure 4.11b. The other two are quickly identifiable due to their possessing of structure dissimilar to noise. Figure H.2a, upon comparison with Figure H.2b, demonstrates that Ile2 is the cause of such structure, and not Lys20. Ile2 is a more extreme residue than Leu5 at the N-terminal end of $\epsilon 2$. Ideally, we would like to use residues close to the ends of the peptides, but Ile2 is clearly erratic. We wish to see the behaviour of the coiled coil as a whole, and Leu5 will be more representative. It will give a slightly lower separation value than if Ile2 was used, but the actual values of the separation are less important than the differences in separation over time; the repressor feet are not included in these test systems, and their size will make the differences here (of the order of a few Å) insignificant in the final system. However, the changes the system can undergo are important for our simulation studies. Figure H.2c demonstrates another spectrum with structure, but Figure H.2d shows that this is due to the inner residue, which was later seen to often form part of the erratic central region. Any residue separations for which this structure was seen were investigated further: if the outer residue was proven to be the cause of an irregular structure, the end-to-end

separations were scrutinised to confirm that no unusual artifacts (such as in H.2a) were seen. Generally, very few problems were found.



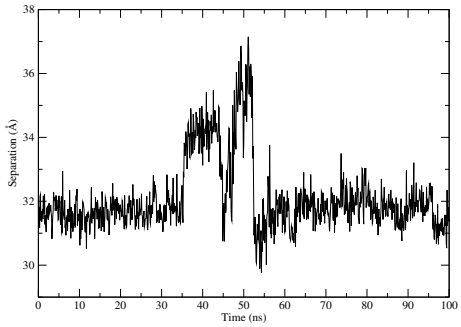
(a) Ile7 - Leu31 (b) Ile7 - Glu27

Figure H.1: Evolution of the coiled coil measurements from Figure 4.11b for the cis-azobenzene system. The same residues as the azobenzene-free system were initially chosen, but Leu31 was found to be part of the uncoiling region of peptide 1, and so separations using Glu27 as its replacement were taken in addition, to ensure system representative behaviour was measured. It was found that the uncoiling had very little effect on the measurements taken; as can be seen here, there is only one siginificant structure in the Ile7-Leu31 separation, and it remains for only a small proportion of the simulation.

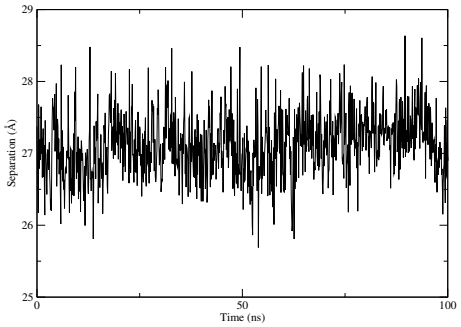


(a) Ile2 - Lys20

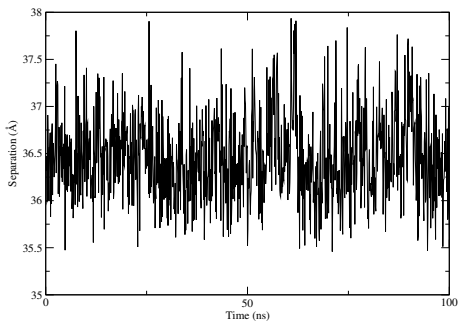
(b) Leu5 - Lys 20



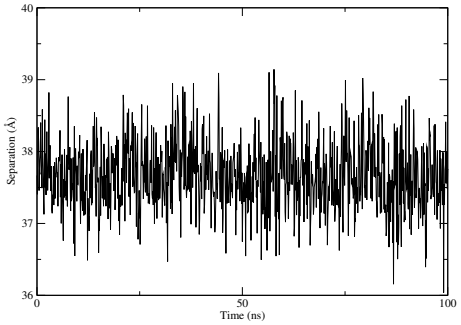
(c) Ile3 - Asn24



(d) Ile3 - Glu21



(e) Ile7 - Leu31



(f) Lys33 - Leu58

Figure H.2: Evolution of the coiled coil measurements from Figure 4.11b for the azobenzene-free system. The chosen residues show a variation of 2 - 3Å for a separation variation that in an ideal system would be zero. The initial choice of Ile2 at the extreme of $\epsilon 2$ was replaced by Leu5 due to Ile2's movement independent of the rest of the coiled coil (indicated in the Ile2-Lys20 separation shown here). The first choice of Asn24 as the inner residue with which to compare Ile3 was replaced by Glu21, due to the erratic region shown in the Ile3-Asn24 separation, found to be due to Asn24 as opposed to the extremely placed residue of interest.

H.2 Vector-based Angle Calculations

The coordinates of two residues (a pair of a or d residues) at both extremities of each of the two coiled coils (8 residues in total, Figure 4.12c, H.4a) were used to calculate midpoints between the two peptides at each of the coiled-coil ends (red points shown in Figure H.4b), and from these, two vectors representing the axes of the coiled coils were constructed (Figure H.4c). The angle between two vectors, if they shared an origin, could then be calculated using Equation H.1 (angle and vectors defined in Figure H.3), but this lack of shared origin is the main issue with the trans system measurements. The closest to a realistic angle determinable reasonably was calculated, by treating the vectors as if they had a shared origin, but remaining in the true directions of their parent coiled coils (Figure H.4d).

$$cos\theta = \frac{\vec{ab} \cdot \vec{cd}}{|ab|\,|cd|} \tag{H.1}$$

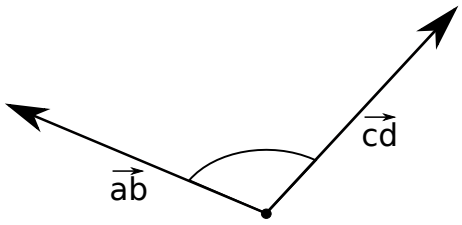
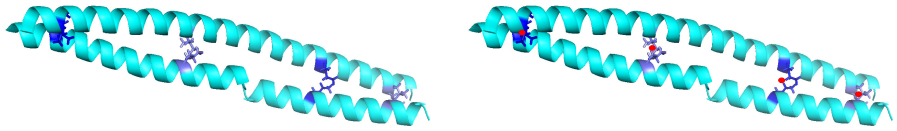
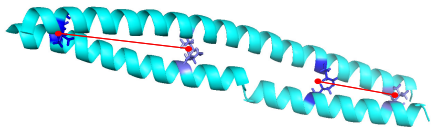


Figure H.3: The angle between two vectors can be calculated using Equation H.1.



(a) Four a and d pairs were chosen

(b) Midpoints were calculated



(c) Vectors were constructed

(d) An angle was calculated for the vectors with a shared origin

Figure H.4: Construction of vectors from residue co-ordinates, to use with Equation H.1.

Bibliography

[1] Bromley, E. H. C. *et al. HFSP Journal* 3:204–212 (2009).

[2] Bromley, E. H. C. *et al. ACS Chemical Biology* 3:38–50 (2008).

[3] Colgin, N., Flinn, T. and Cobb, S. L. *Organic and Biomolecular Chemistry* 9:1864–1870 (2011).

[4] Bath, J., Green, S. J. and Turberfield, A. J. *Angewandte Chemie, International Edition* 44:4358–4361 (2005).

[5] Tian, Y. *et al. Angewandte Chemie, International Edition* 44:4355–4358 (2005).

[6] Shin, J. and Pierce, N. A. *Journal of the American Chemical Society* 126:10834–10835 (2004).

[7] Sherman, W. B. and Seeman, N. C. *Nano Letters* 4:1203–1207 (2004).

[8] Yin, P. *et al. Angewandte Chemie, International Edition* 43:4906–4911 (2004).

[9] Green, S., Bath, J. and Turberfield, A. J. *Physical Review Letters* 101(23):238101 (2008).

[10] Bath, J. *et al. Small* 5(13):1513–6 (2009).

[11] Venkataraman, S. *et al. Nature Nanotechnology* 2(8):490–494 (2007).

[12] Wickham, S. F. J. *et al. Nature Nanotechnology* 6(3):166–9 (2011).

[13] Wickham, S. F. J. *et al. Nature Nanotechnology* 7(3):169–73 (2012).

[14] Gibson, D. G. *et al. Science* 329:52–56 (2010).

[15] Lartigue, C. *et al. Science* 317:632–638 (2007).

[16] Mallik, R. and Gross, S. P. *Current Biology* 14:R971–R982 (2004).

[17] Kodera, N. *et al. Nature* 468(7320):72–6 (2010).

[18] Yildiz, A. *et al. Science* 303(5658):676–8 (2004).

[19] Asbury, C. L. *Current Opinion in Cell Biology* 17(1):89–97 (2005).

[20] Qiu, W. *et al. Nature Structural and Molecular Biology* 19(2):193–200 (2012).

[21] Hammer, J. A. and Sellers, J. R. *Nature Reviews. Molecular Cell Biology* 13(1):13–26 (2012).

[22] De La Cruz, E. M. *et al. Proceedings of the National Academy of Sciences of the United States of America* 96(24):13726–13731 (1999).

[23] Veigel, C. *et al. Nature Cell Biology* 7:861 – 869 (2005).

[24] Sakamoto, T. *et al. Nature* 455(7209):128–132 (2008).

[25] Purcell, T. J., Sweeney, H. L. and Spudich, J. a. *Proceedings of the National Academy of Sciences of the United States of America* 102(39):13873–13878 (2005).

[26] Forgacs, E. *et al. Journal of Biological Chemistry* 283(2):766–773 (2008).

[27] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.3r1 (2010).

[28] Liu, J. *et al. Nature* 442(7099):208–11 (2006).

[29] Ariga, K., Mori, T. and Hill, J. P. *Soft Matter* 8(1):15 (2012).

[30] Watson, J. D. and Crick, F. H. C. *Nature* 171:737–738 (1953).

[31] Nelson, P., Radosavljević, M. and Bromberg, S. *Biological Physics (Updated Edition)*, (W. H. Freeman 2008).

[32] Beves, J. E. *et al. Journal of the American Chemical Society* 136(5):2094–2100 (2014).

[33] De Bo, G. *et al. Journal of the American Chemical Society* 136:5811–5814 (2014).

[34] Feringa, B. L. *Journal of Organic Chemistry* 72(18):6635–52 (2007).

[35] Browne, W. R. *et al. Journal of the American Chemical Society* 128(38):12412–3 (2006).

[36] Bissell, R. *et al. Nature* 369:133–137 (1994).

[37] Shirai, Y. *et al. Nano Letters* 5:2330–2334 (2005).

[38] Shirai, Y. *et al. Chemical Society Reviews* 35(11):1043–55 (2006).

[39] Wyatt, G. R. *Journal of General Physiology* 36(2):201–205 (1952).

[40] Zamenhof, S., Brawerman, G. and Chargaff, E. *Biochimica et Biophysica Acta* 9:402–405 (1952).

[41] Chargaff, E., Lipshitz, R. and Green, C. *Journal of Biological Chemistry* 195:155–160 (1952).

[42] Chargaff, E., Zamenhof, S. and Green, C. *Nature* 165:756–757 (1950).

[43] Saenger, W. *Principles of nucleic acid structure.* Springer advanced texts in chemistry, (Springer-Verlag 1984).

[44] Watson, J. and Crick, F. *Nature* 171:964–967 (1953).

[45] Bath, J. and Turberfield, A. J. *Nature Nanotechnology* 2:275–284 (2007).

[46] http://www.nature.com/nature/journal/v440/n7082/abs/nature04586.html, accessed 1/7/2014.

[47] Douglas, S. M. *et al. Nature* 459(7245):414–8 (2009).

[48] Goodman et al, R. P. *Science* 310:1661–1665 (2005).

[49] Goodman et al, R. P. *Nature Nanotechnology* 3:93–96 (2008).

[50] Erben, C. M., Goodman, R. P. and Turberfield, A. J. *Angewandte Chemie, International Edition* 45:7414–7417 (2006).

[51] Crawford, R. *et al. Angewandte Chemie (International ed. in English)* 52(8):2284–8 (2013).

[52] Walsh, A. S. *et al. ACS Nano* (7):5427–5432 (2011).

[53] Zhang, S. *et al. Proceedings of the National Academy of Sciences of the United States of America* 90(8):3334–8 (1993).

[54] Schneider, J. P. *et al. Journal of the American Chemical Society* 124(50):15030–7 (2002).

[55] Ozbas, B. *et al. Macromolecules* 37(19):7331–7337 (2004).

[56] Pandya, M. J. *et al. Biochemistry* 39:8728–8734 (2000).

[57] Ryadnov, M. G. and Woolfson, D. N. *Nature Materials* 2(5):329–32 (2003).

[58] Papapostolou, D. *et al. Proceedings of the National Academy of Sciences of the United States of America* 104(26):10853–8 (2007).

[59] Villard, V. *et al. Journal of Peptide Science* 12(3):206–12 (2006).

[60] Wang, A. Y. *et al. Biomacromolecules* 9(10):2929–36 (2008).

[61] Pires, M. M. and Chmielewski, J. *Journal of the American Chemical Society* 131(7):2706–12 (2009).

[62] Boyle, A. L. and Woolfson, D. N. *Chemical Society Reviews* 40:4295–4306 (2011).

[63] Salick, D. a. *et al. Journal of the American Chemical Society* 129(47):14793–9 (2007).

[64] Kretsinger, J. K. *et al. Biomaterials* 26(25):5177–86 (2005).

[65] Zhou, M. *et al. Biomaterials* 30(13):2523–30 (2009).

[66] Potekhin, S. a. *et al. Chemistry and Biology* 8(11):1025–32 (2001).

[67] Pires, M. M., Przybyla, D. E. and Chmielewski, J. *Angewandte Chemie (International ed. in English)* 48(42):7813–7 (2009).

[68] Wang, A. Y. *et al. Journal of the American Chemical Society* 127(12):4130–1 (2005).

[69] Silva, G. a. *et al. Science* 303(5662):1352–5 (2004).

[70] Tysseling-Mattiace, V. M. *et al. Journal of Neuroscience* 28(14):3814–23 (2008).

[71] Aulisa, L. *et al. Acta Biomaterialia* 5(3):842–53 (2009).

[72] Fernandez-Lopez, S. *et al. Nature* 412(6845):452–5 (2001).

[73] Horne, W. S. *et al. Bioorganic and Medicinal Chemistry* 13(17):5145–53 (2005).

[74] Motiei, L. *et al. Chemical Communications* (25):3693–5 (2009).

[75] Raman, S. *et al. Nanomedicine : Nanotechnology, Biology, and Medicine* 2(2):95–102 (2006).

[76] Boato, F. *et al. Angewandte Chemie (International ed. in English)* 46(47):9015–8 (2007).

[77] Ambrogelly, A., Palioura, S. and Söll, D. *Nature Chemical Biology* 3:29–35 (2007).

[78] Hudson, A. S. *et al. Tetrahedron Letters* 54(36):4865–4867 (2013).

[79] Branden, C. I. and Tooze, C. *Introduction to Protein Structure*, (Garland Publishing Inc. 1999), (13–35). Second ed.

[80] O'Neil, K. and DeGrado, W. *Science* 250(4981):646–651 (1990).

[81] Horovitz, A., Matthews, J. M. and Fersht, A. R. *Journal of Molecular Biology* 227:560–568 (992).

[82] Chakrabartty, A., Kortemme, T. and Baldwin, R. L. *Protein Science* 3(5):843–52 (1994).

[83] Piela, L., Némethy, G. and Scheraga, H. A. *Biopolymers* 26(9):1587–1600 (1987).

[84] Paoli, M. *et al. Journal of Molecular Biology* 256(4):775–92 (1996).

[85] Murshudov, G. N. *et al. Acta Crystallographica Section D* 58(12):1972–1982 (2002).

[86] Schevitz, R. W. *et al. Nature* 317:782–786 (1985).

[87] Schumacher, M. A. *et al. Science* 266(5186):763–770 (1994).

[88] McKay, D. B. and Steitz, T. A. *Nature* 290:744–749 (1981).

[89] Anderson, W. F. *et al. Nature* 290:754–758 (1981).

[90] Ohlendorf, D. H. *et al. Journal of Molecular Biology* 169(3):757–69 (1983).

[91] Pabo, C. O. and Lewis, M. *Nature* 298:443–447 (1982).

[92] Lehninger, A. *Biochemistry*, (Worth Publishers 1975).

[93] O'Shea, E. K. *et al. Science* 254:539–544 (1991).

[94] Kramer, R. Z. *et al. Nature Structural Biology* 6(5):454–457 (1999).

[95] Lee, M. S. *et al. Science* 245(4918):635–7 (1989).

[96] Albright, R. and Matthews, B. *Journal of Molecular Biology* 280:137–151 (1998).

[97] Crick, F. H. C. *Acta Crystallographica* 6:689–697 (1953).

[98] Pauling, L. and Corey, R. B. *Nature* 171:59–61 (1953).

[99] Lupas, A. *Trends in Biochemical Sciences* 21:375–382 (1996).

[100] Berman, H. *et al. Nucleic Acids Research* 28:235–242 (2000).

[101] Lupas, A., Van Dyke, M. and Stock, J. *Science* 252(5009):1162–1164 (1991).

[102] Berger, B. *et al. Proceedings of the National Academy of Sciences of the United States of America* 92:8259–8263 (1995).

[103] A.V. McDonnell, A. K., T. Jiang and Berger, B. *Bioinformatics* 22:356–358 (2006).

[104] Delorenzi, M. and Speed, T. *Bioinformatics* 18:617–625 (2002).

[105] Fariselli, P. *et al.* In S. Hochreiter and R. Wagner (Eds.), *Bioinformatics Research and Development*, (Springer Berlin Heidelberg2007), vol. 4414 of *Lecture Notes in Computer Science*, (292–302).

[106] Tanizawa, H., Ghimire1, G. D. and Mitaku, S. *Chem-Bio Informatics Journal* 8(3):96–111 (2008).

[107] Altschul, S. F. *et al. Nucleic Acids Research* 25(17):3389–3402 (1997).

[108] Bartoli, L. *et al. Bioinformatics* 25:2757–2763 (2009).

[109] Gruber, M., J, S. and Lupas, A. *Journal of Structural Biology* 155(2):140–145 (2006).

[110] Armstrong, C. T. *et al. Bioinformatics* 27(14):1908–14 (2011).

[111] Wolf, E., Kim, P. S. and Berger, B. *Protein Science* 6(6):1179–1189 (1997).

[112] Rackham, O. *et al. Journal of Molecular Biology* 403(3):480–493 (2010).

[113] Walshaw, J. and Woolfson, D. N. *Journal of Molecular Biology* 307:1427–1450 (2001).

[114] Moutevelis, E. and Woolfson, D. N. *Journal of Molecular Biology* 385:726–732 (2009).

[115] St Maurice, M. *et al. Journal of Biological Chemistry* 282(4):2596–2605 (2007).

[116] Zaccai, N. R. *et al. Nature Chemical Biology* 7:935–941 (2011).

[117] Harbury, P. B. *et al. Science* 262:1401–1407 (1993).

[118] Harbury, P. B., Kim, P. S. and Alber, T. *Nature* 371:80–83 (1994).

[119] Landschulz, W. H., Johnson, P. F. and McKnight, S. L. *Science* 240:1759–1764 (1988).

[120] Stetefeld, J. *et al. Nature Structural and Molecular Biology* 7(9):772 – 776 (2000).

[121] Harbury, P. B. *et al. Science* 282:1462–1467 (1998).

[122] Walshaw, J. and Woolfson, D. N. *Protein Science* 10:668–673 (2001).

[123] Glover, J. N. M. and Harrison, S. C. *Nature* 373:257–261 (1995).

[124] O'Shea, E. K., Lumb, K. J. and Kim, P. S. *Current Biology* 3:658–667 (1993).

[125] Lumb, K. J. and Kim, P. S. *Biochemistry* 34:8642–8648 (1995).

[126] Gonzalez, L. *et al. Nature Structural Biology* 3:1002–1010 (1996).

[127] Gonzalez, L., Woolfson, D. N. and Alber, T. *Nature Structural Biology* 3:1011–1018 (1996).

[128] Woolfson, D. N. and Alber, T. *Protein Science* 4:1596–1607 (1995).

[129] Fletcher, J. M. *et al. ACS Synthetic Biology* 1:240–50 (2012).

[130] Lovejoy, B. *et al. Science* 259:1288–1293 (1993).

[131] Ogihara, N. L. *et al. Protein Science* 6:80–88 (1997).

[132] Channon, K., Bromley, E. H. and Woolfson, D. N. *Current Opinion in Structural Biology* 18:491–498 (2008).

[133] Woolfson, D. N. *Advances in Protein Chemistry* 70:79–112 (2005).

[134] Gribbon, C. *et al. Biochemistry* 47(39):10365–71 (2008).

[135] Bromley, E. H. C. *et al. Journal of the American Chemical Society* 131:928–930 (2009).

[136] Fong, J. H., Keating, A. E. and Singh, M. *Genome Biology* 5(2):R11 (2004).

[137] Mason, J. M. *et al. Proceedings of the National Academy of Sciences of the United States of America* 103(24):8989–8994 (2006).

[138] Boyle, A. L. *et al. Journal of the American Chemical Society* 134(37):15457–67 (2012).

[139] Fletcher, J. M. *et al. Science* 340:595–9 (2013).

[140] Malashkevich, V. *et al. Science* 274(5288):761–765 (1996).

[141] Schroeder, U. *et al. Journal of Molecular Biology* 386(5):1368–81 (2009).

[142] Babapoor, S. *et al. Influenza Research and Treatment* 2011:126794 (2011).

[143] Kaba, S. a. *et al. Journal of Immunology* 183(11):7268–77 (2009).

[144] Pimentel, T. a. P. F. *et al. Chemical Biology and Drug Design* 73(1):53–61 (2009).

[145] Wahome, N. *et al. Chemical Biology and Drug Design* 80(3):349–57 (2012).

[146] Kuwada, N. J., Blab, G. A. and Linke, H. *Chemical Physics* 375(2-3):479–485 (2010).

[147] Niman, C. S. *et al. Lab on a Chip* 13(12):2389–96 (2013).

[148] Kuwada, N. J. *et al. Physical Review E* 84(3):031922 (2011).

[149] Kovacic, S. *et al. Journal of Nanomaterials* 2012:109238 (2012).

[150] Otwinowski, Z. *et al. Nature* 335(6188):321–329 (1988).

[151] Pohl, E., Holmes, R. K. and Hol, W. G. *Journal of Molecular Biology* 292(3):653–667 (1999).

[152] Zhang, R. *et al. Nature* 327(6123):591–597 (1987).

[153] Schiering, N. *et al. Proceedings of the National Academy of Sciences of the United States of America* 92(21):9843–9850 (1995).

[154] Qiu, X. *et al. Structure* 3(1):87–100 (1995).

[155] Boyd, J., Oza, M. N. and Murphy, J. R. *Proceedings of the National Academy of Sciences of the United States of America* 87(15):5968–5972 (1990).

[156] Barksdale, L. *Bacteriological Reviews* 34(4):378–422 (1970).

[157] Pohl, E., Holmes, R. K. and Hol, W. G. J. *Journal of Biological Chemistry* 273(35):22420–22427 (1998).

[158] Schmitt, M. P., Twiddy, E. M. and Holmes, R. K. *Proceedings of the National Academy of Sciences of the United States of America* 89(16):7576–7580 (1992).

[159] Schmitt, M. P. and Holmes, R. K. *Journal of Bacteriology* 176(4):1141–1149 (1994).

[160] Qiu, X. *et al. Biochemistry* 35(38):12292–12302 (1996).

[161] Drawcoil 1.0, http://www.grigoryanlab.org/drawcoil/.

[162] Merrifield, R. B. *Journal of the American Chemical Society* 85:2149–2154 (1963).

[163] Fischer, E. and Fourneau, E. *Berichte der Deutschen Chemischen Gesellschaft* 34(August):2868–2877 (1901).

[164] Fischer, E. *Berichte der Deutschen Chemischen Gesellschaft* 40:1754–1767 (1907).

[165] Goodman, M., Cai, W. and Smith, N. D. *Journal of Peptide Science* 9(9):594–603 (2003).

[166] Yu, H.-M., Chen, S.-T. and Wang, K.-T. *Journal of Organic Chemistry* 57(18):4781–4784 (1992).

[167] Palasek, S. A., Cox, Z. J. and Collins, J. M. *Journal of Peptide Science* 13(3):143–8 (2007).

[168] Friligou, I. *et al. Amino Acids* 40(5):1431–40 (2011).

[169] Bacsa, B., Bosze, S. and Kappe, C. O. *Journal of Organic Chemistry* 75(6):2103–6 (2010).

[170] Bray, B. L. *Nature Reviews. Drug Discovery* 2(July):587–593 (2003).

[171] Coin, I., Beyermann, M. and Bienert, M. *Nature Protocols* 2(12):3247–56 (2007).

[172] Bailey, P. D. *An Introduction to Peptide Chemistry*, (Wiley , Chichester 1990).

[173] *Hydrophobic Interaction and Reversed Phase Chromatography: Principles and Methods*, (GE Healthcare, Elanders Tofters 2006).

[174] Karas, M. and Hillenkamp, F. *Analytical Chemistry* 60(20):2299–2301 (1988).

[175] Duncan, M. W., Roder, H. and Hunsucker, S. W. *Briefings in Functional Genomics and Proteomics* 7(5):355–70 (2008).

[176] Hillenkamp, F. *et al. Analytical Chemistry* 63(24):1193A–1203A (1991).

[177] Trauger, S. A., Webb, W. and Siuzdak, G. *Spectroscopy* 16(1):15–28 (2002).

[178] Sathyanarayana, D. *Electronic Absorption Spectroscopy and Related Techniques*, (Universities Press 2001).

[179] Rodger, A. and Nordén, B. *Circular Dichroism and Linear Dichroism.* Oxford Chemistry Masters, (Oxford University Press 1997).

[180] Greenfield, N. and Fasman, G. D. *Biochemistry* 8(10):4108–16 (1969).

[181] Benevides, J. M. *et al. Biochemistry* 27(3):931–8 (1988).

[182] Greenfield, N. J. *Analytical Biochemistry* 235:1–10 (1996).

[183] Provencher, S. W. and Glöckner, J. *Biochemistry* 20(1):33–7 (1981).

[184] Sreerama, N. and Woody, R. *Analytical Biochemistry* 209:32–44 (1993).

[185] Sreerama, N. and Woody, R. W. *Analytical Biochemistry* 287(2):252–60 (2000).

[186] Johnson, W. C. *Proteins* 35(3):307–12 (1999).

[187] Louis-Jeune, C., Andrade-Navarro, M. A. and Perez-Iratxeta, C. *Proteins* 80(2):374–381 (2012).

[188] Brasseur, G. and Solomon, S. *Aeronomy of the Middle Atmosphere: Chemistry and Physics of the Stratosphere and Mesosphere.* Atmospheric and Oceanographic Sciences Library, (Springer 2006).

[189] Pecora, R. *Dynamic Light Scattering: Applications of Photon Correlation Spectroscopy*, (Springer 1985).

[190] Berne, B. and Pecora, R. *Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics.* Dover Books on Physics Series, (Dover Publications 2000).

[191] Cotterill, R. M. J. *Biophysics: An Introduction*, (John Wiley & Sons 2002).

[192] Sheehan, D. *Physical Biochemistry: Principles and Applications*, (John Wiley & Sons 2009).

[193] Scott, D. J. and Schuck, P. In D. Scott, S. Harding and A. Rowe (Eds.), *Analytical Ultracentrifugation: Techniques and Methods*, (RSC, Cambridge 2005), chap. 1, (1–25).

[194] Winzor, D. and Harding, S. In S. Harding and B. Chowdhry (Eds.), *Protein-Ligand Interactions: Hydrodynamics and Calorimetry*, (OUP, Oxford 2001), chap. 5, (105–135). 1 ed.

[195] Futaki, S. and Kitagawa, K. *Tetrahedron Letters* 35(8):1267–1270 (1994).

[196] Flint, D. G. *et al. Chemistry and Biology* 9(02):391–397 (2002).

[197] Kumita, J. R., Smart, O. S. and Woolley, G. A. *Proceedings of the National Academy of Sciences of the United States of America* 97(8):3803–8 (2000).

[198] Kumita, J. R. *et al. Protein Engineering Design and Selection* 15(7):561–569 (2002).

[199] Renner, C. and Moroder, L. *ChemBioChem* 7(6):868–878 (2006).

[200] Muñoz, V. and Serrano, L. *Nature Structural and Molecular Biology* 1(6):399–409 (1994).

[201] Muoz, V. and Serrano, L. *Journal of Molecular Biology* 245(3):297 – 308 (1995).

[202] Muñoz, V. and Serrano, L. *Journal of Molecular Biology* 245:275–296 (1995).

[203] Muñoz, V. and Serrano, L. *Biopolymers* 41(5):495–509 (1997).

[204] Lacroix, E., Viguera, A. R. and Serrano, L. *Journal of Molecular Biology* 284(1):173 – 191 (1998).

[205] Zimm, B. H. and Bragg, J. K. *Journal of Chemical Physics* 31(2):526–535 (1959).

[206] Lifson, S. and Roig, a. *Journal of Chemical Physics* 34(6):1963–1974 (1961).

[207] Mason, J. M. and Arndt, K. M. *Chembiochem : a European journal of chemical biology* 5(2):170–6 (2004).

[208] Alder, B. J. and Wainwright, T. E. *Journal of Chemical Physics* 27(5):1208 (1957).

[209] Berendsen, H., van der Spoel, D. and van Drunen, R. *Computer Physics Communications* 91(1-3):43–56 (1995).

[210] Lindahl, E., Hess, B. and van der Spoel, D. *Journal of Molecular Modeling* 7:306–317 (2001).

[211] Van Der Spoel, D. *et al. Journal of Computational Chemistry* 26(16):1701–18 (2005).

[212] Hess, B. *et al. Journal of Chemical Theory and Computation* 4:435–447 (2008).

[213] Pronk, S. *et al. Bioinformatics* 29(7):845–54 (2013).

[214] Cornell, W. D. *et al. Journal of the American Chemical Society* 117(19):5179–5197 (1995).

[215] Lindorff-Larsen, K. *et al. PLoS ONE* 7(2):e32131 (2012).

[216] Koukos, P. I. and Glykos, N. M. *The Journal of Physical Chemistry B* 118(34):10076–10084. PMID: 25098230 (2014).

[217] Koukos, P. I. and Glykos, N. M. *Journal of Molecular Graphics and Modelling* 41:68–71 (2013).

[218] Alder, B. J., Frankel, S. P. and Lewinson, V. A. *Journal of Chemical Physics* 23(3):417 (1955).

[219] Cheung, D. L. G. *Structures and Properties of Liquid Crystals and Related Molecules from Computer Simulation.* Ph.D. thesis (2002).

[220] Abraham, M. *et al. www.gromacs.org* (2014).

[221] Humphrey, W., Dalke, A. and Schulten, K. *Journal of Molecular Graphics* 14:33–38 (1996).

[222] Pettersen, E. F. *et al. Journal of Computational Chemistry* 25(13):1605–1612 (2004).